(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
10 September 2004 (10.09.2004)

PCT

(10) International Publication Number
WO 2004/076638 A2

(51) International Patent Classification[7]: C12N

(21) International Application Number:
PCT/US2004/005654

(22) International Filing Date: 25 February 2004 (25.02.2004)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
10/374,780 25 February 2003 (25.02.2003) US
10/675,852 30 September 2003 (30.09.2003) US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier application:
US 10/374,780 (CIP)
Filed on 25 February 2003 (25.02.2003)

(71) Applicant (*for all designated States except US*): **MENDEL BIOTECHNOLOGY, INC.** [US/US]; 21375 Cabot Boulevard, Hayward, CA 94545 (US).

(72) Inventors; and

(75) Inventors/Applicants (*for US only*): **SHERMAN, Bradley, K.** [US/US]; 1039 Overlook Road, Berkeley, CA 94708 (US). **RIECHMANN, Jose, Luis** [ES/US]; 518 S. El Molino Avenue, #308, Pasadena, CA 91101 (US). **RATCLIFFE, Oliver** [GB/US]; 814 East 21st Street, Oakland, CA 94606 (US). **JIANG, Cai-Zhong** [US/US]; 34495 Heathrow Terrace, Fremont, CA 94555 (US). **HEARD, Jacqueline, E.** [US/US]; 21 Whitaker Drive, Stonington, CT 06378 (US). **HAAKE, Volker** [DE/DE]; Lichterfelder Ring 206, 12209 Berlin (DE). **CREELMAN, Robert, A.** [US/US]; 2801 Jennifer Drive, Castro Valley, CA 94546 (US). **ADAM, Luc, J.** [CA/US]; 25800 Industrial Boulevard, Apt. 1403, Hayward, CA 94545 (US). **REUBER, Lynne, T.** [US/US]; 1115 S. Grant Street, San Mateo, CA 94402 (US). **KEDDIE, James, S.** [GB/US]; 20-14th Avenue, San Mateo, CA 94402 (US). **DUBELL, Arnold, N.** [US/US]; 14857 Wake Ave., San Leandro, CA 94578 (US). **PINEDA, Omaira** [CO/US]; 4060 9th Place, Vero Beach, FL 32960 (US). **REPETTI, Peter, P.** [US/US]; 1200 65th Street, Apt. 231, Berkley, CA 94709-1820, Emeryville, CA 94608 (US). **CENTURY, Karen, S.** [US/US]; 741 Jackson Street, Albany, CA 94706 (US). **GUTTERSON, Neal, I.** [US/US]; 5169 Golden Gate Ave., Oakland, CA 94618 (US). **YU, Guo-Liang** [US/US]; 242 Gravatt Drive, Berkeley, CA 94705-1531 (US). **BROUN, Pierre, E.** [FR/GB]; Department of Biology (Area 7), University of York, P.O. Box 373, York, Yorkshire Y01 05YW (GB). **KUMIMOTO, Roderick, W.** [US/US]; 517 Oak Ave., San Bruno, CA 94066 (US). **PILGRIM, Marsha, L.** [US/US]; 1368 Patrick Henry Drive, Phoenixville, PA 19460 (US).

(74) Agents: **WARD, Michael** et al.; Morrison & Foerster, LLP, 425 Market Street, San Francisco, CA 94105-2482 (US).

(81) Designated States (*unless otherwise indicated, for every kind of national protection available*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States (*unless otherwise indicated, for every kind of regional protection available*): ARIPO (BW, GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Declaration under Rule 4.17:
— *of inventorship (Rule 4.17(iv)) for US only*

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: POLYNUCLEOTIDES AND POLYPEPTIDES IN PLANTS

(57) Abstract: The invention relates to plant transcription factor polypeptides, polynucleotides that encode them, homologs from a variety of plant species, and methods of using the polynucleotides and polypeptides to produce transgenic plants having advantageous properties compared to a reference plant. Sequence information related to these polynucleotides and polypeptides can also be used in bioinformatic search methods is also disclosed.

WO 2004/076638 A2

POLYNUCLEOTIDES AND POLYPEPTIDES IN PLANTS

RELATIONSHIP TO COPENDING APPLICATIONS

This application claims the benefit of U.S. Non-provisional Application No. 10/374,780, filed 25 February, 2003, and U.S. Non-provisional Application No. 10/675,852, filed 30 September, 2003, the contents of which are hereby incorporated by reference in their entirety.

TECHNICAL FIELD

This invention relates to the field of plant biology. More particularly, the present invention pertains to compositions and methods for modifying a plant phenotypically.

BACKGROUND OF THE INVENTION

A plant's traits, such as its biochemical, developmental, or phenotypic characteristics, may be controlled through a number of cellular processes. One important way to manipulate that control is through transcription factors - proteins that influence the expression of a particular gene or sets of genes. Transformed and transgenic plants that comprise cells having altered levels of at least one selected transcription factor, for example, possess advantageous or desirable traits. Strategies for manipulating traits by altering a plant cell's transcription factor content can therefore result in plants and crops with new and/or improved commercially valuable properties.

Transcription factors can modulate gene expression, either increasing or decreasing (inducing or repressing) the rate of transcription. This modulation results in differential levels of gene expression at various developmental stages, in different tissues and cell types, and in response to different exogenous (e.g., environmental) and endogenous stimuli throughout the life cycle of the organism.

Because transcription factors are key controlling elements of biological pathways, altering the expression levels of one or more transcription factors can change entire biological pathways in an organism. For example, manipulation of the levels of selected transcription factors may result in increased expression of economically useful proteins or biomolecules in plants or improvement in other agriculturally relevant characteristics. Conversely, blocked or reduced expression of a transcription factor may reduce biosynthesis of unwanted compounds or remove an undesirable trait. Therefore, manipulating transcription factor levels in a plant offers tremendous potential in agricultural biotechnology for modifying a plant's traits. A number of the agriculturally relevant characteristics of plants, and desirable traits that may be imbued by gene expression are listed below.

Useful Plant Traits

Category: abiotic stress; desired trait: chilling tolerance

The term "chilling sensitivity" has been used to describe many types of physiological damage produced at low, but above freezing, temperatures. Most crops of tropical origins such as soybean, rice, maize and cotton are easily damaged by chilling. Typical chilling damage includes wilting, necrosis, chlorosis or leakage of ions from cell membranes. The underlying mechanisms of chilling sensitivity are not completely understood yet, but probably involve the level of membrane saturation and other physiological deficiencies. For example, photoinhibition of photosynthesis (disruption of photosynthesis due to high light intensities) often occurs under clear atmospheric conditions subsequent to cold late summer/autumn nights. By some estimates, chilling accounts for monetary losses in the United States (US) second only to drought and flooding. For example, chilling may lead to yield losses and lower product quality through the delayed ripening of maize. Another consequence of poor growth is the rather poor ground cover of maize fields in spring, often resulting in soil erosion, increased occurrence of weeds, and reduced uptake of nutrients. A retarded uptake of mineral nitrogen could also lead to increased losses of nitrate into the ground water.

Category: abiotic stress; desired trait: freezing tolerance.

Freezing is a major environmental stress that limits where crops can be grown and reduces yields considerably, depending on the weather in a particular growing season. In addition to exceptionally stressful years that cause measurable losses of billions of dollars, less extreme stress almost certainly causes smaller yield reductions over larger areas to produce yield reductions of similar dollar value every year. For instance, in the US, the 1995 early fall frosts are estimated to have caused losses of over one billion dollars to corn and soybeans. The spring of 1998 saw an estimated $200 M of damages to Georgia alone, in the peach, blueberry and strawberry industries. The occasional freezes in Florida have shifted the citrus belt further south due to $100 M or more losses. California sustained $650 M of damage in 1998 to the citrus crop due to a winter freeze. In addition, certain crops such as *Eucalyptus*, which has the very favorable properties of rapid growth and good wood quality for pulping, are not able to grow in the southeastern states due to occasional freezes.

Inherent winter hardiness of the crop determines in which agricultural areas it can survive the winter. For example, for wheat, the northern central portion of the US has winters that are too cold for good winter wheat crops. Approximately 20% of the US wheat crop is spring wheat, with a market value of $2 billion. Areas growing spring wheat could benefit by growing winter wheat that had increased winter hardiness. Assuming a 25% yield increase when growing winter wheat, this would create $500 M of increased value. Additionally, the existing winter wheat is severely stressed by freezing conditions and should have improved yields with increased tolerance to these stresses. An estimate of the yield benefit of these traits is 10% of the $4.4 billion winter wheat crop in the US or $444 M of yield increase, as well as better survival in extreme freezing conditions that occur periodically.

Thus plants more resistant to freezing, both midwinter freezing and sudden freezes, would protect a farmers' investment, improve yield and quality, and allow some geographies to grow more profitable and productive crops. Additionally, winter crops such as canola, wheat and barley have 25% to 50% yield increases relative to spring planted varieties of the same crops. This yield increase is due to the "head start" the fall planted crop has over the spring planted crop and its reaching maturity earlier while the temperatures, soil moisture and lack of pathogens provide more favorable conditions.

Category: abiotic stress; desired trait: salt tolerance.

One in five hectares of irrigated land is damaged by salt, an important historical factor in the decline of ancient agrarian societies. This condition is only expected to worsen, further reducing the availability of arable land and crop production, since none of the top five food crops - wheat, corn, rice, potatoes, and soybean - can tolerate excessive salt.

Detrimental effects of salt on plants are a consequence of both water deficit resulting in osmotic stress (similar to drought stress) and the effects of excess sodium ions on critical biochemical processes. As with freezing and drought, high saline causes water deficit; the presence of high salt makes it difficult for plant roots to extract water from their environment (Buchanan et al. (2000) in Biochemistry and Molecular Biology of Plants, American Society of Plant Physiologists, Rockville, MD). Soil salinity is thus one of the more important variables that determines where a plant may thrive. In many parts of the world, sizable land areas are uncultivable due to naturally high soil salinity. To compound the problem, salination of soils that are used for agricultural production is a significant and increasing problem in regions that rely heavily on agriculture. The latter is compounded by over-utilization, over-fertilization and water shortage, typically caused by climatic change and the demands of increasing population. Salt tolerance is of particular importance early in a plant's lifecycle, since evaporation from the soil surface causes upward water movement, and salt accumulates in the upper soil layer where the seeds are placed. Thus, germination normally takes place at a salt concentration much higher than the mean salt level in the whole soil profile.

Category: abiotic stress; desired trait: drought tolerance.

While much of the weather that we experience is brief and short-lived, drought is a more gradual phenomenon, slowly taking hold of an area and tightening its grip with time. In severe cases, drought can last for many years, and can have devastating effects on agriculture and water supplies. With burgeoning population and chronic shortage of available fresh water, drought is not only the number one weather related problem in agriculture, it also ranks as one of the major natural disasters of all time, causing not only economic damage, but also loss of human lives. For example, losses from the US drought of 1988 exceeded $40 billion, exceeding the losses caused by Hurricane Andrew in 1992, the Mississippi River floods of 1993, and the San Francisco earthquake in 1989. In some areas of the world, the effects of

drought can be far more severe. In the Horn of Africa the 1984–1985 drought led to a famine that killed 750,000 people.

Problems for plants caused by low water availability include mechanical stresses caused by the withdrawal of cellular water. Drought also causes plants to become more susceptible to various diseases (Simpson (1981). "The Value of Physiological Knowledge of Water Stress in Plants", In Water Stress on Plants, (Simpson, G. M., ed.), Praeger, NY, pp. 235-265).

In addition to the many land regions of the world that are too arid for most if not all crop plants, overuse and over-utilization of available water is resulting in an increasing loss of agriculturally-usable land, a process which, in the extreme, results in desertification. The problem is further compounded by increasing salt accumulation in soils, as described above, which adds to the loss of available water in soils.

Category: abiotic stress; desired trait: heat tolerance.

Germination of many crops is very sensitive to temperature. A transcription factor that would enhance germination in hot conditions would be useful for crops that are planted late in the season or in hot climates.

Seedlings and mature plants that are exposed to excess heat may experience heat shock, which may arise in various organs, including leaves and particularly fruit, when transpiration is insufficient to overcome heat stress. Heat also damages cellular structures, including organelles and cytoskeleton, and impairs membrane function (Buchanan, *supra*).

Heat shock may result a decrease in overall protein synthesis, accompanied by expression of heat shock proteins. Heat shock proteins function as chaperones and are involved in refolding proteins denatured by heat.

Category: abiotic stress; desired trait: tolerance to low nitrogen and phosphorus.

The ability of all plants to remove nutrients from their environment is essential to survival. Thus, identification of genes that encode polypeptides with transcription factor activity may allow for the generation of transgenic plants that are better able to make use of available nutrients in nutrient-poor environments.

Among the most important macronutrients for plant growth that have the largest impact on crop yield are nitrogenous and phosphorus-containing compounds. Nitrogen- and phosphorus- containing fertilizers are used intensively in agriculture practices today. An increase in grain crop yields from 0.5 to 1.0 metric tons per hectare to 7 metric tons per hectare accompanied the use of commercial fixed nitrogen fertilizer in production farming (Vance (2001) *Plant Physiol.* 127: 390-397). Given current practices, in order to meet food production demands in years to come, considerable increases in the amount of nitrogen- and phosphorus-containing fertilizers will be required (Vance, *supra*).

Nitrogen is the most abundant element in the Earth's atmosphere yet it is one of the most limiting elements to plant growth due to its lack of availability in the soil. Plants obtain N from the soil from several sources including commercial fertilizers, manure and the mineralization of organic matter. The intensive use of N fertilizers in present agricultural practices is problematic, the energy intensive Haber-Bosch process makes N fertilizer and it is estimated that the US uses annually between 3-5% of the nation's natural gas for this process. In addition to the expense of N fertilizer production and the depletion of non-renewable resources, the use of N fertilizers has led to the eutrophication of freshwater ecosystems and the contamination of drinking water due to the runoff of excess fertilizer into ground water supplies.

Phosphorus is second only to N in its importance as a macronutrient for plant growth and to its impact on crop yield. Phosphorus (P) is extremely immobile and not readily available to roots in the soil and is therefore often growth limiting to plants. Inorganic phosphate (Pi) is a constituent of several important molecules required for energy transfer, metabolic regulation and protein activation (Marschner (1995) Mineral Nutrition of Higher Plants, 2nd ed., Academic Press, San Diego, CA). Plants have evolved several strategies to help cope with P and N deprivation that include metabolic as well as developmental adaptations. Most, if not all, of these strategies have components that are regulated at the level of transcription and therefore are amenable to manipulation by transcription factors. Metabolic adaptations include increasing the availability of P and N by increasing uptake from the soil though the induction of high affinity and low affinity transporters, and/or increasing its mobilization in the plant. Developmental adaptations include increases in primary and secondary roots, increases in root hair number and length, and associations with mycorrhizal fungi (Bates and Lynch (1996) *Plant Cell Environ.* 19: 529-538; Harrison (1999) *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 50: 361-389).

Category: biotic stress: desired trait: disease resistance.

Disease management is a significant expense in crop production worldwide. According to EPA reports for 1996 and 1997, US farmers spend approximately $6 billion on fungicides annually. Despite this expenditure, according to a survey conducted by the food and agriculture organization, plant diseases still reduce worldwide crop productivity by 12% and in the United States alone, economic losses due to plant pathogens amounts to 9.1 billion dollars (FAO, 1993). Data from these reports and others demonstrate that despite the availability of chemical control only a small proportion of the losses due to disease can be prevented. Not only are fungicides and anti-bacterial treatments expensive to growers, but their widespread application poses both environmental and health risks. The use of plant biotechnology to engineer disease resistant crops has the potential to make a significant economic impact on agriculture and forestry industries in two ways: reducing the monetary and environmental expense of fungicide application and reducing both pre-harvest and post-harvest crop losses that occur now despite the use of costly disease management practices.

Fungal, bacterial, oomycete, viral, and nematode diseases of plants are ubiquitous and important problems, and often severely impact yield and quality of crop and other plants. A very few examples of diseases of plants include:

Powdery mildew, caused by the fungi *Erysiphe, Sphaerotheca, Phyllactinia, Microsphaera, Podosphaera, or Uncinula*, in, for example, wheat, bean, cucurbit, lettuce, pea, grape, tree fruit crops, as well as roses, phlox, lilacs, grasses, and Euonymus;

Fusarium-caused diseases such as *Fusarium* wilt in cucurbits, *Fusarium* head blight in barley and wheat, wilt and crown and root rot in tomatoes;

Sudden oak death, caused by the oomycete *Phytophthora ramorum*; this disease was first detected in 1995 in California tan oaks. The disease has since killed more than 100,000 tan oaks, coast live oaks, black oaks, and Shreve's oaks in coastal regions of northern California, and more recently in southwestern Oregon (Roach (2001) *National Geographic News*, December 6, 2001);

Black Sigatoka, a fungal disease caused by *Mycosphaerella* species that attacks banana foliage, is spreading throughout the regions of the world that are responsible for producing most of the world's banana crop;

Eutypa dieback, caused by *Eutypa lata*, affects a number of crop plants, including vine grape. Eutypa dieback delays shoot emergence, and causes chlorosis, stunting, and tattering of leaves;

Pierce's disease, caused by the bacterium *Xylella fastidiosa*, precludes growth of grapes in the southeastern United States, and threatens the profitable wine grape industry in northern California. The bacterium clogs the vasculature of the grapevines, resulting in foliar scorching followed by slow death of the vines. There is no known treatment for Pierce's disease;

Bacterial Spot caused by the bacterium *Xanthomonas campestris* causes serious disease problems on tomatoes and peppers. It is a significant problem in the Florida tomato industry because it spreads rapidly, especially in warm periods where there is wind-driven rain. Under these conditions, there are no adequate control measures;

Diseases caused by viruses of the family Geminiviridae are a growing agricultural problem worldwide. Geminiviruses have caused severe crop losses in tomato, cassava, and cotton. For instance, in the 1991-1992 growing season in Florida, geminiviruses caused $140 million in damages to the tomato crop (Moffat (1991) *Science* 286: 1835). Geminiviruses have the ability to recombine between strains to rapidly produce new virulent varieties. Therefore, there is a pressing need for broad-spectrum geminivirus control;

The soybean cyst nematode, *Heterodera glycines*, causes stunting and chlorosis of soybean plants, which results in yield losses or plant death from severe infestation. Annual losses in the United States have been estimated at $1.5 billion (University of Minnesota Extension Service).

The aforementioned pathogens represent a very small fraction of diverse species that seriously affect plant health and yield. For a more complete description of numerous plant diseases, see, for example, Vidhyasekaran (1997) Fungal Pathogenesis in Plants and Crops: Molecular Biology and Host

Defense Mechanisms, Marcel Dekker, Monticello, NY), or Agrios (1997) Plant Pathology, Academic Press, New York, NY). Plants that are able to resist disease may produce significantly higher yields and improved food quality. It is thus of considerable importance to find genes that reduce or prevent disease.

Category: light response; desired trait: reduced shade avoidance.

Shade avoidance describes the process in which plants grown in close proximity attempt to outcompete each other by increasing stem length at the expense of leaf, fruit and storage organ development. This is caused by the plant's response to far-red radiation reflected from leaves of neighboring plants, which is mediated by phytochrome photoreceptors. Close proximity to other plants, as is produced in high-density crop plantings, increases the relative proportion of far-red irradiation, and therefore induces the shade avoidance response. Shade avoidance adversely affects biomass and yield, particularly when leaves, fruits or other storage organs constitute the desired crop (see, for example, Smith (1982) *Annu. Rev. Plant Physiol.* 33: 481-518; Ballare et al. (1990) *Science* 247: 329-332; Smith (1995) *Annu. Dev. Plant Physiol. Mol. Biol.*, 46: 289-315; and Schmitt et al. (1995), *American Naturalist*, 146: 937-953). Alteration of the shade avoidance response in tobacco through alteration of phytochrome levels has been shown to produce an increase in harvest index (leaf biomass/total biomass) at high planting density, which would result in higher yield (Robson et al. (1996) *Nature Biotechnol.* 14: 995-998).

Category: flowering time; desired trait: altered flowering time and flowering control.

Timing of flowering has a significant impact on production of agricultural products. For example, varieties with different flowering responses to environmental cues are necessary to adapt crops to different production regions or systems. Such a range of varieties have been developed for many crops, including wheat, corn, soybean, and strawberry. Improved methods for alteration of flowering time will facilitate the development of new, geographically adapted varieties.

Breeding programs for the development of new varieties can be limited by the seed-to-seed cycle. Thus, breeding new varieties of plants with multi-year cycles (such as biennials, e.g. carrot, or fruit trees, such as citrus) can be very slow. With respect to breeding programs, there would be a significant advantage in having commercially valuable plants that exhibit controllable and modified periods to flowering ("flowering times"). For example, accelerated flowering would shorten crop and tree breeding programs.

Improved flowering control allows more than one planting and harvest of a crop to be made within a single season. Early flowering would also improve the time to harvest plants in which the flower portion of the plant constitutes the product (e.g., broccoli, cauliflower, and other edible flowers). In addition, chemical control of flowering through induction or inhibition of flowering in plants could provide a significant advantage to growers by inducing more uniform fruit production (e.g., in strawberry)

A sizable number of plants for which the vegetative portion of the plant forms the valuable crop tend to "bolt" dramatically (e.g., spinach, onions, lettuce), after which biomass production declines and product quality diminishes (e.g., through flowering-triggered senescence of vegetative parts). Delay or prevention of flowering may also reduce or preclude dissemination of pollen from transgenic plants.

Category: growth rate; desired trait: modified growth rate.

For almost all commercial crops, it is desirable to use plants that establish more quickly, since seedlings and young plants are particularly susceptible to stress conditions such as salinity or disease. Since many weeds may outgrow young crops or out-compete them for nutrients, it would also be desirable to determine means for allowing young crop plants to out compete weed species. Increasing seedling growth rate (emergence) contributes to seedling vigor and allows for crops to be planted earlier in the season with less concern for losses due to environmental factors. Early planting helps add days to the critical grain-filling period and increases yield.

Providing means to speed up or slow down plant growth would also be desirable to ornamental horticulture. If such means be provided, slow growing plants may exhibit prolonged pollen-producing or fruiting period, thus improving fertilization or extending harvesting season.

Category: growth rate; desired trait: modified senescence and cell death.

Premature senescence, triggered by various plant stresses, can limit production of both leaf biomass and seed yield. Transcription factor genes that suppress premature senescence or cell death in response to stresses can provide means for increasing yield. Delay of normal developmental senescence could also enhance yield, particularly for those plants for which the vegetative part of the plant represents the commercial product (e.g., spinach, lettuce).

Although leaf senescence is thought to be an evolutionary adaptation to recycle nutrients, the ability to control senescence in an agricultural setting has significant value. For example, a delay in leaf senescence in some maize hybrids is associated with a significant increase in yields and a delay of a few days in the senescence of soybean plants can have a large impact on yield. In an experimental setting, tobacco plants engineered to inhibit leaf senescence had a longer photosynthetic lifespan, and produced a 50% increase in dry weight and seed yield (Gan and Amasino (1995) *Science* 270: 1986-1988). Delayed flower senescence may generate plants that retain their blossoms longer and this may be of potential interest to the ornamental horticulture industry, and delayed foliar and fruit senescence could improve post-harvest shelf-life of produce.

Further, programmed cell death plays a role in other plant responses, including the resistance response to disease, and some symptoms of diseases, for example, as caused by necrotrophic pathogens such as *Botrytis cinerea* and *Sclerotinia sclerotiorum* (Dickman et al. *Proc. Natl. Acad. Sci.*, 98: 6957-6962). Localized senescence and/or cell death can be used by plants to contain the spread of harmful microorganisms. A specific localized cell death response, the "hypersensitive response", is a component

of race-specific disease resistance mediated by plant resistance genes. The hypersensitive response is thought to help limit pathogen growth and to initiate a signal transduction pathway that leads to the induction of systemic plant defenses. Accelerated senescence may be a defense against obligate pathogens, such as powdery mildew, that rely on healthy plant tissue for nutrients. With regard to powdery mildew, *Botrytis cinerea* and *Sclerotinia sclerotiorum* and other pathogens, transcription factors that ameliorate cell death and/or damage may reduce the significant economic losses encountered, such as, for example, *Botrytis cinerea* in strawberry and grape.

Category: growth regulator; desired trait: altered sugar sensing

Sugars are key regulatory molecules that affect diverse processes in higher plants including germination, growth, flowering, senescence, sugar metabolism and photosynthesis. Sucrose, for example, is the major transport form of photosynthate and its flux through cells has been shown to affect gene expression and alter storage compound accumulation in seeds (source-sink relationships). Glucose-specific hexose-sensing has also been described in plants and is implicated in cell division and repression of "famine" genes (photosynthetic or glyoxylate cycles).

Category: morphology; desired trait: altered morphology

Trichomes are branched or unbranched epidermal outgrowths or hair structures on a plant. Trichomes produce a variety of secondary biochemicals such as diterpenes and waxes, the former being important as, for example, insect pheromones, and the latter as protectants against desiccation and herbivorous pests. Since diterpenes also have commercial value as flavors, aromas, pesticides and cosmetics, and potential value as anti-tumor agents and inflammation-mediating substances, they have been both products and the target of considerable research. In most cases where the metabolic pathways are impossible to engineer, increasing trichome density or size on leaves may be the only way to increase plant productivity. Thus, it would be advantageous to discover trichome-affecting transcription factor genes for the purpose of increasing trichome density, size, or type to produce plants that are better protected from insects or that yield higher amounts of secondary metabolites.

The ability to manipulate wax composition, amount, or distribution could modify plant tolerance to drought and low humidity or resistance to insects, as well as plant appearance. In particular, a possible application for a transcription factor gene that reduces wax production in sunflower seed coats would be to reduce fouling during seed oil processing. Antisense or co-suppression of transcription factors involved in wax biosynthesis in a tissue specific manner can be used to specifically alter wax composition, amount, or distribution in those plants and crops from which wax is either a valuable attribute or product or an undesirable constituent of plants.

Other morphological characteristics that may be desirable in plants include those of an ornamental nature. These include changes in seed color, overall color, leaf and flower shape, leaf color, leaf size, or glossiness of leaves. Plants that produce dark leaves may have benefits for human health;

flavonoids, for example, have been used to inhibit tumor growth, prevent of bone loss, and prevention lipid oxidation in animals and humans. Plants in which leaf size is increased would likely provide greater biomass, which would be particularly valuable for crops in which the vegetative portion of the plant constitutes the product. Plants with glossy leaves generally produce greater epidermal wax, which, if it could be augmented, resulted in a pleasing appearance for many ornamentals, help prevent desiccation, and resist herbivorous insects and disease-causing agents. Changes in plant or plant part coloration, brought about by modifying, for example, anthocyanin levels, would provide novel morphological features.

In many instances, the seeds of a plant constitute a valuable crop. These include, for example, the seeds of many legumes, nuts and grains. The discovery of means for producing larger seed would provide significant value by bringing about an increase in crop yield.

Plants with altered inflorescence, including, for example, larger flowers or distinctive floral configurations, may have high value in the ornamental horticulture industry.

Modifications to flower structure may have advantageous or deleterious effects on fertility, and could be used, for example, to decrease fertility by the absence, reduction or screening of reproductive components. This could be a desirable trait, as it could be exploited to prevent or minimize the escape of the pollen of genetically modified organisms into the environment.

Manipulation of inflorescence branching patterns may also be used to influence yield and offer the potential for more effective harvesting techniques. For example, a "self pruning" mutation of tomato results in a determinate growth pattern and facilitates mechanical harvesting (Pnueli et al. (2001) *Plant Cell* 13(12): 2687-2702).

Alterations of apical dominance or plant architecture could create new plant varieties. Dwarf plants may be of potential interest to the ornamental horticulture industry.

Category: seed biochemistry; desired trait: altered seed oil

The composition of seeds, particularly with respect to seed oil quantity and/or composition, is very important for the nutritional value and production of various food and feed products. Desirable improvements to oils include enhanced heat stability, improved nutritional quality through, for example, reducing the number of calories in seed, increasing the number of calories in animal feeds, or altering the ratio of saturated to unsaturated lipids comprising the oils.

Category: seed biochemistry; desired trait: altered seed protein

As with seed oils, seed protein content and composition is very important for the nutritional value and production of various food and feed products. Altered protein content or concentration in seeds may be used to provide nutritional benefits, and may also prolong storage capacity, increase seed pest or disease resistance, or modify germination rates. Altered amino acid composition of seeds, through altered protein composition, is also a desired objective for nutritional improvement.

Category: seed biochemistry; desired trait: altered prenyl lipids.

Prenyl lipids, including the tocopherols, play a role in anchoring proteins in membranes or membranous organelles. Tocopherols have both anti-oxidant and vitamin E activity. Modified tocopherol composition of plants may thus be useful in improving membrane integrity and function, which may mitigate abiotic stresses such as heat stress. Increasing the anti-oxidant and vitamin content of plants through increased tocopherol content can provide useful human health benefits.

Category: leaf biochemistry; desired trait: altered glucosinolate levels

Increases or decreases in specific glucosinolates or total glucosinolate content can be desirable depending upon the particular application. For example: (i) glucosinolates are undesirable components of the oilseeds used in animal feed, since they produce toxic effects; low-glucosinolate varieties of canola have been developed to combat this problem; (ii) some glucosinolates have anti-cancer activity; thus, increasing the levels or composition of these compounds can be of use in production of nutraceuticals; and (iii) glucosinolates form part of a plant's natural defense against insects; modification of glucosinolate composition or quantity could therefore afford increased protection from herbivores. Furthermore, tissue specific promoters can be used in edible crops to ensure that these compounds accumulate specifically in particular tissues, such as the epidermis, which are not taken for human consumption.

Category: leaf biochemistry; desired trait: flavonoid production.

Expression of transcription factors that increase flavonoid production in plants, including anthocyanins and condensed tannins, may be used to alter pigment production for horticultural purposes, and possibly to increase stress resistance. Flavonoids have antimicrobial activity and could be used to engineer pathogen resistance. Several flavonoid compounds have human health promoting effects such as inhibition of tumor growth, prevention of bone loss and prevention of lipid oxidation. Increased levels of condensed tannins in forage legumes would provide agronomic benefits in ruminants by preventing pasture bloat by collapsing protein foams within the rumen. For a review on the utilities of flavonoids and their derivatives, see Dixon et al. (1999) *Trends Plant Sci.* 4: 394-400.

The present invention relates to methods and compositions for producing transgenic plants with modified traits, particularly traits that address the agricultural and food needs described in the above background information. These traits may provide significant value in that they allow the plant to thrive in hostile environments, where, for example, temperature, water and nutrient availability or salinity may limit or prevent growth of non-transgenic plants. The traits may also comprise desirable morphological alterations, larger or smaller size, disease and pest resistance, alterations in flowering time, light response, and others.

We have identified polynucleotides encoding transcription factors, developed numerous transgenic plants using these polynucleotides, and have analyzed the plants for a variety of important traits. In so doing, we have identified important polynucleotide and polypeptide sequences for producing commercially valuable plants and crops as well as the methods for making them and using them. Other aspects and embodiments of the invention are described below and can be derived from the teachings of this disclosure as a whole.

SUMMARY OF THE INVENTION

The present invention pertains to transgenic plants that comprise a recombinant polynucleotide that includes a nucleotide sequence encoding a HAP3, HAP3-like or HAP5 CCAAT transcription factor with the ability to regulate abiotic stress tolerance in a plant.

Transgenic plants and methods for producing transgenic plants are provided. The transgenic plants comprise a recombinant polynucleotide having a polynucleotide sequence, or a sequence that is complementary to this polynucleotide sequence, that encodes a transcription factor.

The present invention is directed to transgenic plants with altered expression of a transcription factor polypeptide. When these plants are compared to non-transgenic plants or wild-type control plants of the same species (that is, those that do not have the expression of the transcription factor polypeptide altered), the transgenic plants are shown to have increased tolerance to a variety of abiotic stresses. These may include, for example, osmotic stress, drought, salt stress, heat stress, or cold stress. A sizeable number of transcription factor sequences, which may be found in the Sequence Listing, have been used to confer these advantages. For example, the transgenic plant may comprise as part of its genome a transgene encoding a polypeptide member of the CCAAT-box binding transcription factor family or the MYB-related transcription factor family. By overexpressing these polypeptide members, the plant becomes more abiotic stress tolerant wild-type controls.

In a further refinement of the invention, the polypeptide member may be of the CCAAT box-binding transcription factor family, and more specifically, from the G482 subclade of the non-LEC1-like clade of proteins of the L1L-related CCAAT transcription factor family. These polypeptides comprise a B domain (B domains are further characterized in the specification that follows, with examples from various plant species listed). The B domains of these polypeptides are able to bind a regulatory region of DNA comprising the motif CCAAT, and this binding is responsible for the regulation of transcription of that DNA. This regulation of transcription confers increased abiotic stress tolerance in the transgenic plant.

Alternatively, the polypeptide member may be a MYB-related transcription factor, and more specifically a member of the G682 subclade of MYB-related transcription factors. These polypeptides comprise a MYB-related domain (MYB-related domains further characterized in the specification that follows, with examples from various plant species listed). The MYB-related domains of these polypeptides are able to bind a regulatory region of DNA, and this binding is responsible for the

regulation of transcription of that DNA. Similar to the regulation of transcription by CCAAT family transcription factors, increased abiotic stress tolerance in the transgenic plant is achieved by the binding of the MYB-related transcription factor to the DNA.

In another alternative, the polypeptide member may be a member of a wide variety of transcription factor families and groups, exemplified by the transcription factor polypeptides of the invention that have been shown to confer useful advantages or traits in plants when the expression of these polypeptides is altered in the plants.

The transgenic plants of the invention may be either dicotyledonous or monocotyledonous, and the polynucleotide and polypeptide sequences of the invention may be derived from dicotyledonous or monocotyledonous plants, or be creations that are structurally and functionally similar. The transgenic plants that define the invention may be crossed with other plants or themselves to generate progeny plants, which may possess advantageous characteristics such as improved stress tolerance. Seed derived from the transgenic plants of the invention are also encompassed within the scope of the invention.

The invention is also directed to methods for producing a transgenic plant having increased tolerance to abiotic stress by introducing an expression vector that contains a polynucleotide sequence encoding a polypeptide of the invention into the plant.

The polypeptide may comprise, for example, a member of G482 subclade of the non-LEC1-like clade of proteins of the L1L-related CCAAT transcription factor family, or a member of the G682 subclade of MYB-related transcription factors, or a member of any of the other transcription factor families or groups of the invention that have been shown to provide useful traits in plants when their expression is altered. In the case where the polypeptide comprises a G482 subclade member, the polypeptide will comprise a B domain that is of sufficiently homology to the B domain of G481, SEQ ID NO: 88 that the B domain functions in the same manner as the G481 domain by binding to DNA at a transcription regulating region comprising the motif CCAAT, thus regulating transcription of the DNA and confers increased abiotic stress tolerance in the transgenic plant as compared to non-transgenic plants of the same species.

Alternatively, the expression vector may also contain a polynucleotide sequence that encodes a polypeptide of the G682 subclade of MYB-related transcription factors, the MYB-related domain of the polypeptide being sufficiently homologous to the MYB-related domain of G682, SEQ ID NO: 148,that the MYB-related domain binds to DNA at a transcription regulating region and brings about transcriptional and increased abiotic stress tolerance in the transgenic plant.

Regardless of the nucleotide sequence used, the expression vector will also contain one or more regulatory elements operably linked to the nucleotide sequence. It is through the regulatory elements that expression of the nucleotide sequence is controlled in a target plant. By introducing the expression vector into a cell of the target plant, growing the plant cell into a plant and allowing the plant to overexpress the polypeptide, a plant with improved abiotic stress tolerance (relative to, for example, wild-type controls) is generated. The improved plants may be identified by comparison to with one or more control plants.

Any plant with increased stress tolerance produced by this method may be crossed with itself or another plant. Seed that develops as a result of this crossing may then be selected, and a progeny plant grown from the seed. This would have the effect of producing a transgenic progeny plant that would have increased tolerance to abiotic stress, as compared to a non-transgenic plant of the same species.

The invention also pertains to a method for increasing a plant's tolerance to abiotic stress by providing a vector comprising a nucleotide of the invention or a nucleotide encoding a polypeptide of the invention, and regulatory elements operably linked to the nucleotide sequence. These regulatory elements are able to control expression of the nucleotide sequence in a target plant. The target plant is transformed with the vector to generate a transformed plant with increased tolerance to abiotic stress compared to non-transgenic plants of the same species. The sequences used in this method may include, for example, polynucleotide or polypeptide members of the G482 subclade of the non-LEC1-like clade of proteins of the L1L-related CCAAT transcription factor family, or of the G682 subclade of MYB-related transcription factors. In the case of either of these examples, the conserved domain that is, the B or MYB-related domain, would be sufficiently homologous to the corresponding domains of SEQ ID NO: 88 or 148, G481 or G682, respectively, that the conserved domain binds to DNA at a transcription regulating region, and thus regulates transcription and increases abiotic stress tolerance in the plant as compared to a non-transformed or transgenic plant of the same species.

Brief Description of the Sequence Listing and Drawings

The Sequence Listing provides exemplary polynucleotide and polypeptide sequences of the invention. The traits associated with the use of the sequences are included in the Examples.

CD-ROM1 (Copy 1) is a read-only memory computer-readable compact disc and contains a copy of the Sequence Listing in ASCII text format. The Sequence Listing is named "MBI0047PCT.ST25.txt" and is 6,312 kilobytes in size. The copies of the Sequence Listing on the CD-ROM disc are hereby incorporated by reference in their entirety.

CD-ROM2 (Copy 2) is an exact copy of CD-R1 (Copy 1).

CD-ROM3 contains a computer-readable format (CRF) copy of the Sequence Listing as a text (.txt) file.

Figure 1 shows a conservative estimate of phylogenetic relationships among the orders of flowering plants (modified from Angiosperm Phylogeny Group (1998) *Ann. Missouri Bot. Gard.* 84: 1-49). Those plants with a single cotyledon (monocots) are a monophyletic clade nested within at least two major lineages of dicots; the eudicots are further divided into rosids and asterids. *Arabidopsis* is a rosid

eudicot classified within the order Brassicales; rice is a member of the monocot order Poales. Figure 1 was adapted from Daly et al. (2001) *Plant Physiol.* 127: 1328-1333.

Figure 2 shows a phylogenic dendogram depicting phylogenetic relationships of higher plant taxa, including clades containing tomato and *Arabidopsis*; adapted from Ku et al. (2000) *Proc. Natl. Acad. Sci.* 97: 9121-9126; and Chase et al. (1993) *Ann. Missouri Bot. Gard.* 80: 528-580.

Figures 3A, and 3B show an alignment of G682 (SEQ ID NO: 148) and polynucleotide sequences that are paralogous and orthologous to G682. The alignment was produced using MACVECTOR software (Accelrys, Inc., San Diego, CA).

Figures 4A, 4B, 4C and 4D show an alignment of G867 (SEQ ID NO: 170) and polynucleotide sequences that are paralogous and orthologous to G867. The alignment was produced using MACVECTOR software (Accelrys, Inc.).

Figures 5A, 5B, 5C, 5D, 5E and 5F show an alignment of G912 (SEQ ID NO: 186) and polynucleotide sequences that are paralogous and orthologous to G912. The alignment was produced using MACVECTOR software (Accelrys, Inc.).

Figure 6 is adapted from Kwong et al (2003) Plant Cell 15: 5-18, and shows crop sequences identified through BLAST analysis of various L1L-related sequences. A phylogeny tree was then generated using ClustalX based on whole protein sequences showing the non-LEC1-like HAP3 clade of transcription factors (large box). This clade contains *Arabidopsis* sequences as well as members from diverse species that are phylogenetically distinct from the LEC1-like proteins (seen in the next two figures). The smaller box delineates the G482-like subclade, containing transcription factors that are structurally most closely related to G481 and G482.

Similar to Figure 6, Figure 7 shows the phylogenic relationship of sequences within the G482-subclade (within the smaller box) and the non-LEC1-like clade (larger box). The G482-subclade contains members from diverse species, including *Arabidopsis*, soy, rice and corn sequences that are phylogenetically distinct from the LEC1-like proteins, some of which are also shown in Figure 6. *Arabidopsis* plants carrying a mutation in the LEC1 gene display embryos that are intolerant to desiccation and that show defects in seed maturation (Lotan et al. (1998) *Cell* 93: 1195-1205). We have shown that the G482 subclade members confer abiotic stress tolerance. Given their phylogenetic divergence, it is possible that LEC1-like proteins seen below the G482 subclade have evolved to confer desiccation tolerance specifically within the embryo, whereas other G482 subclade evolved to confer abiotic stress tolerance in non-embryonic tissues.

For Figure 8, a phylogenic relationship of sequences within the G482-subclade was generated with different method than used to create Figure 7. In this case, the phylogenetic tree and multiple sequence alignments of G481 and related full length proteins were constructed using ClustalW (CLUSTAL W Multiple Sequence Alignment Program version 1.83, 2003) and MEGA2 (http://www.megasoftware.net) software. The ClustalW multiple alignment parameters were:

Gap Opening Penalty :10.00

Gap Extension Penalty :0.20

Delay divergent sequences :30 %

DNA Transitions Weight :0.50

Protein weight matrix :Gonnet series

DNA weight matrix :IUB

Use negative matrix :OFF.

A FastA formatted alignment was then used to generate a phylogenetic tree in MEGA2 using the neighbor joining algorithm and a p-distance model. A test of phylogeny was done via bootstrap with 100 replications and Random Speed set to default. Cut off values of the bootstrap tree were set to 50%. The G482 subclade of the non-LEC1-like clade of proteins of the L1L-related CCAAT transcription factor family (box), are derived from a common single node (arrow) and a representative number, including *Arabidopsis* sequences G481, G482, G485, G1364 and G2345, soy sequences G3470, G3471, G3472, G3475, G3476, rice sequences G3395, G3397, G3398, G3429, and corn sequences G3434, and G3436, have been shown to confer abiotic stress tolerance in plants (see Example XIII).

Figure 9 shows the domain structure of HAP3 proteins. HAP3 proteins contain an amino-terminal A domain, a central B domain, and a carboxy-terminal C domain. There may be relatively little sequence similarity between HAP3 proteins in the A and C domains. The A and C domains could thus provide a degree of specificity to each member of the HAP3 family. The B domain is the conserved region that specifies DNA binding and subunit association.

In Figures 10A-10F, the alignments of HAP3 polypeptides are presented, including G481, G482, G485, G1364, G2345, G1781 and related sequences from *Arabidopsis* aligned with soybean, rice and corn sequences, showing the B domains (indicated by the line that spans Figures 10B through 10D). Consensus residues within the listed sequences are indicated by boldface. The boldfaced residues in the consensus sequence that appears at the bottom of Figures 10A through 10C in their respective positions are uniquely found in the non-LEC1-like clade. The underlined serine residue appearing in the consensus sequence in its respective positions is uniquely found within the G482-like subclade. As discussed in greater detail below in Example III, the residue positions indicated by the arrows in Figure 10B are associated with an alteration of flowering time when these polypeptides are overexpressed.

Figure 11 is a phylogenetic tree of defined conserved domains of G682 and related polypeptides constructed with ClustalW (CLUSTAL W Multiple Sequence Alignment Program version 1.83, 2003) and MEGA2 (http://www.megasoftware.net) software. ClustalW multiple alignment parameters were as follows:

Gap Opening Penalty :10.00

Gap Extension Penalty :0.20

Delay divergent sequences :30 %

DNA Transitions Weight :0.50

Protein weight matrix :Gonnet series

DNA weight matrix :IUB

Use negative matrix :OFF

A FastA formatted alignment was then used to generate a phylogenetic tree in MEGA2 using the neighbor joining algorithm and a p-distance model. A test of phylogeny was done via bootstrap with 100 replications and Random Speed set to default. Cut off values of the bootstrap tree were set to 50%. The G682 subclade of MYB-related transcription factors, a group of structurally and functionally related sequences that derive from a single ancestral node (arrow), appears within the box in Figure 11.

Figures 12A-12D show the effects of water deprivation and recovery from this treatment on *Arabidopsis* control and 35S::G481-overexpressing lines. After eight days of drought treatment overexpressing plants had a darker green and less withered appearance (Figure 12C) than those in the control group (Figure 12A). The differences in appearance between the control and G481-overexpressing plants after they were rewatered was even more striking. Most (11 of 12 plants; exemplified in Figure 12B) of this set of control plants died after rewatering, indicating the inability to recover following severe water deprivation, whereas all nine of the overexpressor plants of the line shown recovered from this drought treatment (exemplified in Figure 12D). The results shown in Figures 12A-12D were typical of a number of control and 35S::G481-overexpressing lines.

Figures 13A and 13B show the effects of salt stress on *Arabidopsis* seed germination. The three lines of G481-and G482 overexpressors on these two plate had longer roots and showed greater cotyledon expansion (arrows) after three days on 150 mM NaCl than the control seedlings on the right-hand sides of the plates.

In Figure 14A, G481 null mutant seedlings (labeled K481) show reduced tolerance of osmotic stress, relative to the control seedlings in Figure 14B, as evidenced by the reduced cotyledon expansion and root growth in the former group. Without salt stress tolerance on control media, (Figures 14C, G481 null mutants; and 14D, control seedlings), the knocked out and control plants appear the same.

Figures 15A-15D show the effects of stress-related treatments on G485 overexpressing seedlings (35S::G485 lines) in plate assays. In each treatment, including cold, high sucrose, high salt and ABA germination assays, the overexpressors fared much better than the wild-type controls exposed to the same treatments in Figures 15E-15H, respectively, as evidenced by the enhanced cotyledon expansion and root growth seen with the overexpressing seedlings.

Figures 16A - 16C depict the effects of G485 knockout and overexpression on flowering time and maturation. As seen in Figure 16A, a T-DNA insertion knockout mutation containing a SALK_062245 insertion was shown to flower several days later than wild-type control plants. The plants in Figure 16A are shown 44 days after germination. Figure 16C shows that G485 primary transformants flowered distinctly earlier than wild-type controls. These plants are shown 24 days after germination. These effects were observed in each of two independent T1 plantings derived from separate transformation dates. Additionally, accelerated flowering was also seen in plants that overexpressed G485 from a two component system (35S::LexA;op-LexA::G485). These studies indicated that G485 is

both sufficient to act as a floral activator, and is also necessary in that role within the plant. G485 overexpressor plants also matured and set siliques much more rapidly than wild type controls, as shown in Figure 16B with plants 39 days post-germination.

Figure 17A shows the effects of a high salt medium on the germination of 35S::G3392 (a rice G682 ortholog) line 322. In this figure, the seedlings appeared larger and greener than the wild-type *Arabidopsis* seedlings similarly treated in Figure 17B.

Figure 18A shows the effects of cold treatment on the germination of 35S::G3450 (soy G682 ortholog) line 307. In this figure, the seedlings had less anthocyanin than the wild-type *Arabidopsis* seedlings similarly treated in Figure 18B.

Figure 19A compares the germination of 35S::G3431 (a corn G682 ortholog) line 327 *Arabidopsis* seedlings on a 9.4% sucrose germination plate with non-transgenic control seedlings of the same species similarly treated in Figure 19B. The overexpressors were greener and had less anthocyanin than the non-transgenic control plants of the same species.

In Figure 20, the deleterious effects of a heat treatment are seen in the wild-type plants on the right, and to a much lesser extent in the plants overexpressing the soy G3450 sequence on the left.

DETAILED DESCRIPTION OF EXEMPLARY EMBODIMENTS

In an important aspect, the present invention relates to polynucleotides and polypeptides, for example, for modifying phenotypes of plants. Throughout this disclosure, various information sources are referred to and/or are specifically incorporated. The information sources include scientific journal articles, patent documents, textbooks, and World Wide Web browser-inactive page addresses, for example. While the reference to these information sources clearly indicates that they can be used by one of skill in the art, each and every one of the information sources cited herein are specifically incorporated in their entirety, whether or not a specific mention of "incorporation by reference" is noted. The contents and teachings of each and every one of the information sources can be relied on and used to make and use embodiments of the invention.

As used herein and in the appended claims, the singular forms "a," "an," and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to "a plant" includes a plurality of such plants, and a reference to "a stress" is a reference to one or more stresses and equivalents thereof known to those skilled in the art, and so forth.

The polynucleotide sequences of the invention encode polypeptides that are members of well-known transcription factor families, including plant transcription factor families, as disclosed in Tables 7 - 11. Generally, the transcription factors encoded by the present sequences are involved in cellular metabolism, cell differentiation and proliferation and the regulation of growth. Accordingly, one skilled in the art would recognize that by expressing the present sequences in a plant, one may change the expression of autologous genes or induce the expression of introduced genes. By affecting the expression of similar autologous sequences in a plant that have the biological activity of the present sequences, or by

introducing the present sequences into a plant, one may alter a plant's phenotype to one with improved traits. The sequences of the invention may also be used to transform a plant and introduce desirable traits not found in the wild-type cultivar or strain. Plants may then be selected for those that produce the most desirable degree of over- or under-expression of target genes of interest and coincident trait improvement.

The sequences of the present invention may be from any species, particularly plant species, in a naturally occurring form or from any source whether natural, synthetic, semi-synthetic or recombinant. The sequences of the invention may also include fragments of the present amino acid sequences. In this context, a "fragment" refers to a fragment of a polypeptide sequence which is at least 5 to about 15 amino acids in length, most preferably at least 14 amino acids, and which retain some biological activity of a transcription factor. Where "amino acid sequence" is recited to refer to an amino acid sequence of a naturally occurring protein molecule, "amino acid sequence" and like terms are not meant to limit the amino acid sequence to the complete native amino acid sequence associated with the recited protein molecule.

As one of ordinary skill in the art recognizes, transcription factors can be identified by the presence of a region or domain of structural similarity or identity to a specific consensus sequence or the presence of a specific consensus DNA-binding site or DNA-binding site motif (see, for example, Riechmann et al. (2000) *Science* 290: 2105-2110). The plant transcription factors may belong to one of the following transcription factor families: the AP2 (APETALA2) domain transcription factor family (Riechmann and Meyerowitz (1998) *Biol. Chem.* 379: 633-646); the MYB transcription factor family (ENBib; Martin and Paz-Ares (1997) *Trends Genet.* 13: 67-73); the MADS domain transcription factor family (Riechmann and Meyerowitz (1997) *Biol. Chem.* 378: 1079-1101); the WRKY protein family (Ishiguro and Nakamura (1994) *Mol. Gen. Genet.* 244: 563-571); the ankyrin-repeat protein family (Zhang et al. (1992) *Plant Cell* 4: 1575-1588); the zinc finger protein (Z) family (Klug and Schwabe (1995) *FASEB J.* 9: 597-604); Takatsuji (1998) *Cell. Mol. Life Sci.* 54:582-596); the homeobox (HB) protein family (Buerglin (1994) in Guidebook to the Homeobox Genes, Duboule (ed.) Oxford University Press); the CAAT-element binding proteins (Forsburg and Guarente (1989) *Genes Dev.* 3: 1166-1178); the squamosa promoter binding proteins (SPB) (Klein et al. (1996) *Mol. Gen. Genet.* 1996 250: 7-16); the NAM protein family (Souer et al. (1996) *Cell* 85: 159-170); the IAA/AUX proteins (Abel et al. (1995) *J. Mol. Biol.* 251: 533-549); the HLH/MYC protein family (Littlewood et al. (1994) *Prot. Profile* 1: 639-709); the DNA-binding protein (DBP) family (Tucker et al. (1994) *EMBO J.* 13: 2994-3002); the bZIP family of transcription factors (Foster et al. (1994) *FASEB J.* 8: 192-200); the Box P-binding protein (the BPF-1) family (da Costa e Silva et al. (1993) *Plant J.* 4: 125-135); the high mobility group (HMG) family (Bustin and Reeves (1996) *Prog. Nucl. Acids Res. Mol. Biol.* 54: 35-100); the scarecrow (SCR) family (Di Laurenzio et al. (1996) *Cell* 86: 423-433); the GF14 family (Wu et al. (1997) *Plant Physiol.* 114: 1421-1431); the polycomb (PCOMB) family (Goodrich et al. (1997) *Nature* 386: 44-51); the teosinte branched (TEO) family (Luo et al. (1996) *Nature* 383: 794-799); the ABI3 family (Giraudat et

al. (1992) *Plant Cell* 4: 1251-1261); the triple helix (TH) family (Dehesh et al. (1990) *Science* 250: 1397-1399); the EIL family (Chao et al. (1997) *Cell* 89: 1133-44); the AT-HOOK family (Reeves and Nissen (1990) *J. Biol. Chem.* 265: 8573-8582); the S1FA family (Zhou et al. (1995) *Nucleic Acids Res.* 23: 1165-1169); the bZIPT2 family (Lu and Ferl (1995) *Plant Physiol.* 109: 723); the YABBY family (Bowman et al. (1999) *Development* 126: 2387-96); the PAZ family (Bohmert et al. (1998) *EMBO J.* 17: 170-80); a family of miscellaneous (MISC) transcription factors including the DPBF family (Kim et al. (1997) *Plant J.* 11: 1237-1251) and the SPF1 family (Ishiguro and Nakamura (1994) *Mol. Gen. Genet.* 244: 563-571); the GARP family (Hall et al. (1998) *Plant Cell* 10: 925-936), the TUBBY family (Boggin et al (1999) *Science* 286 : 2119-2125), the heat shock family (Wu (1995) *Annu. Rev. Cell Dev. Biol.* 11: 441-469), the ENBP family (Christiansen et al. (1996) *Plant Mol. Biol.* 32: 809-821), the RING-zinc family (Jensen et al. (1998) *FEBS Letters* 436: 283-287), the PDBP family (Janik et al. (1989) *Virology* 168: 320-329), the PCF family (Cubas et al. *Plant J.* (1999) 18: 215-22), the SRS (SHI-related) family (Fridborg et al. (1999) *Plant Cell* 11: 1019-1032), the CPP (cysteine-rich polycomb-like) family (Cvitanich et al. (2000) *Proc. Natl. Acad. Sci.* 97: 8163-8168), the ARF (auxin response factor) family (Ulmasov et al. (1999) *Proc. Natl. Acad. Sci.* 96: 5844-5849), the SWI/SNF family (Collingwood et al. (1999) *J. Mol. Endocrinol.* 23: 255-275), the ACBF family (Seguin et al. (1997) *Plant Mol. Biol.* 35: 281-291), PCGL (CG-1 like) family (da Costa e Silva et al. (1994) *Plant Mol. Biol.* 25: 921-924) the ARID family (Vazquez et al. (1999) *Development* 126: 733-742), the Jumonji family (Balciunas et al. (2000), *Trends Biochem. Sci.* 25: 274-276), the bZIP-NIN family (Schauser et al. (1999) *Nature* 402: 191-195), the E2F family (Kaelin et al. (1992) *Cell* 70: 351-364) and the GRF-like family (Knaap et al. (2000) *Plant Physiol.* 122: 695-704). As indicated by any part of the list above and as known in the art, transcription factors have been sometimes categorized by class, family, and sub-family according to their structural content and consensus DNA-binding site motif, for example. Many of the classes and many of the families and sub-families are listed here. However, the inclusion of one sub-family and not another, or the inclusion of one family and not another, does not mean that the invention does not encompass polynucleotides or polypeptides of a certain family or sub-family. The list provided here is merely an example of the types of transcription factors and the knowledge available concerning the consensus sequences and consensus DNA-binding site motifs that help define them as known to those of skill in the art (each of the references noted above are specifically incorporated herein by reference). A transcription factor may include, but is not limited to, any polypeptide that can activate or repress transcription of a single gene or a number of genes. This polypeptide group includes, but is not limited to, DNA-binding proteins, DNA-binding protein binding proteins, protein kinases, protein phosphatases, protein methyltransferases, GTP-binding proteins, and receptors, and the like.

In addition to methods for modifying a plant phenotype by employing one or more polynucleotides and polypeptides of the invention described herein, the polynucleotides and polypeptides of the invention have a variety of additional uses. These uses include their use in the recombinant production (i.e., expression) of proteins; as regulators of plant gene expression, as diagnostic probes for

the presence of complementary or partially complementary nucleic acids (including for detection of natural coding nucleic acids); as substrates for further reactions, e.g., mutation reactions, PCR reactions, or the like; as substrates for cloning e.g., including digestion or ligation reactions; and for identifying exogenous or endogenous modulators of the transcription factors. A "polynucleotide" is a nucleic acid molecule comprising a plurality of polymerized nucleotides, e.g., at least about 15 consecutive polymerized nucleotides, optionally at least about 30 consecutive nucleotides, at least about 50 consecutive nucleotides. A polynucleotide may be a nucleic acid, oligonucleotide, nucleotide, or any fragment thereof. In many instances, a polynucleotide comprises a nucleotide sequence encoding a polypeptide (or protein) or a domain or fragment thereof. Additionally, the polynucleotide may comprise a promoter, an intron, an enhancer region, a polyadenylation site, a translation initiation site, 5' or 3' untranslated regions, a reporter gene, a selectable marker, or the like. The polynucleotide can be single stranded or double stranded DNA or RNA. The polynucleotide optionally comprises modified bases or a modified backbone. The polynucleotide can be, e.g., genomic DNA or RNA, a transcript (such as an mRNA), a cDNA, a PCR product, a cloned DNA, a synthetic DNA or RNA, or the like. The polynucleotide can be combined with carbohydrate, lipids, protein, or other materials to perform a particular activity such as transformation or form a useful composition such as a peptide nucleic acid (PNA). The polynucleotide can comprise a sequence in either sense or antisense orientations. "Oligonucleotide" is substantially equivalent to the terms amplimer, primer, oligomer, element, target, and probe and is preferably single stranded.

Definitions

A "recombinant polynucleotide" is a polynucleotide that is not in its native state, e.g., the polynucleotide comprises a nucleotide sequence not found in nature, or the polynucleotide is in a context other than that in which it is naturally found, e.g., separated from nucleotide sequences with which it typically is in proximity in nature, or adjacent (or contiguous with) nucleotide sequences with which it typically is not in proximity. For example, the sequence at issue can be cloned into a vector, or otherwise recombined with one or more additional nucleic acid.

An "isolated polynucleotide" is a polynucleotide whether naturally occurring or recombinant, that is present outside the cell in which it is typically found in nature, whether purified or not. Optionally, an isolated polynucleotide is subject to one or more enrichment or purification procedures, e.g., cell lysis, extraction, centrifugation, precipitation, or the like.

A "polypeptide" is an amino acid sequence comprising a plurality of consecutive polymerized amino acid residues e.g., at least about 15 consecutive polymerized amino acid residues, optionally at least about 30 consecutive polymerized amino acid residues, at least about 50 consecutive polymerized amino acid residues. In many instances, a polypeptide comprises a polymerized amino acid residue sequence that is a transcription factor or a domain or portion or fragment thereof. A transcription factor can regulate gene expression and may increase or decrease gene expression in a plant. Additionally, the

polypeptide may comprise 1) a localization domain, 2) an activation domain, 3) a repression domain, 4) an oligomerization domain, or 5) a DNA-binding domain, or the like. The polypeptide optionally comprises modified amino acid residues, naturally occurring amino acid residues not encoded by a codon, non-naturally occurring amino acid residues.

A "recombinant polypeptide" is a polypeptide produced by translation of a recombinant polynucleotide. A "synthetic polypeptide" is a polypeptide created by consecutive polymerization of isolated amino acid residues using methods well known in the art. An "isolated polypeptide," whether a naturally occurring or a recombinant polypeptide, is more enriched in (or out of) a cell than the polypeptide in its natural state in a wild-type cell, e.g., more than about 5% enriched, more than about 10% enriched, or more than about 20%, or more than about 50%, or more, enriched, i.e., alternatively denoted: 105%, 110%, 120%, 150% or more, enriched relative to wild type standardized at 100%. Such an enrichment is not the result of a natural response of a wild-type plant. Alternatively, or additionally, the isolated polypeptide is separated from other cellular components with which it is typically associated, e.g., by any of the various protein purification methods herein.

"Identity" or "similarity" refers to sequence similarity between two polynucleotide sequences or between two polypeptide sequences, with identity being a more strict comparison. The phrases "percent identity" and "% identity" refer to the percentage of sequence similarity found in a comparison of two or more polynucleotide sequences or two or more polypeptide sequences. "Sequence similarity" refers to the percent similarity in base pair sequence (as determined by any suitable method) between two or more polynucleotide sequences. Two or more sequences can be anywhere from 0-100% similar, or any integer value therebetween. Identity or similarity can be determined by comparing a position in each sequence that may be aligned for purposes of comparison. When a position in the compared sequence is occupied by the same nucleotide base or amino acid, then the molecules are identical at that position. A degree of similarity or identity between polynucleotide sequences is a function of the number of identical or matching nucleotides at positions shared by the polynucleotide sequences. A degree of identity of polypeptide sequences is a function of the number of identical amino acids at positions shared by the polypeptide sequences. A degree of homology or similarity of polypeptide sequences is a function of the number of amino acids at positions shared by the polypeptide sequences.

"Alignment" refers to a number of DNA or amino acid sequences aligned by lengthwise comparison so that components in common (i.e., nucleotide bases or amino acid residues) may be visually and readily identified. The fraction or percentage of components in common is related to the homology or identity between the sequences. Alignments such as those of Figures 3A and B, 4A-D, 5A-F and 9A-F may be used to identify conserved domains and relatedness within these domains. An alignment may suitably be determined by means of computer programs known in the art, such as MACVECTOR software (1999) (Accelrys, Inc., San Diego, CA).

The terms "highly stringent" or "highly stringent condition" refer to conditions that permit hybridization of DNA strands whose sequences are highly complementary, wherein these same

conditions exclude hybridization of significantly mismatched DNAs. Polynucleotide sequences capable of hybridizing under stringent conditions with the polynucleotides of the present invention may be, for example, variants of the disclosed polynucleotide sequences, including allelic or splice variants, or sequences that encode orthologs or paralogs of presently disclosed polypeptides. Nucleic acid hybridization methods are disclosed in detail by Kashima et al. (1985) *Nature* 313:402-404, and Sambrook et al. (1989) Molecular Cloning: A Laboratory Manual, 2nd Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y ("Sambrook"); and by Haymes et al., "Nucleic Acid Hybridization: A Practical Approach", IRL Press, Washington, D.C. (1985), which references are incorporated herein by reference.

In general, stringency is determined by the temperature, ionic strength, and concentration of denaturing agents (e.g., formamide) used in a hybridization and washing procedure (for a more detailed description of establishing and determining stringency, see below). The degree to which two nucleic acids hybridize under various conditions of stringency is correlated with the extent of their similarity. Thus, similar nucleic acid sequences from a variety of sources, such as within a plant's genome (as in the case of paralogs) or from another plant (as in the case of orthologs) that may perform similar functions can be isolated on the basis of their ability to hybridize with known transcription factor sequences. Numerous variations are possible in the conditions and means by which nucleic acid hybridization can be performed to isolate transcription factor sequences having similarity to transcription factor sequences known in the art and are not limited to those explicitly disclosed herein. Such an approach may be used to isolate polynucleotide sequences having various degrees of similarity with disclosed transcription factor sequences, such as, for example, transcription factors having 60% identity, or more preferably greater than about 70% identity, most preferably 72% or greater identity with disclosed transcription factors.

The term "equivalog" describes members of a set of homologous proteins that are conserved with respect to function since their last common ancestor. Related proteins are grouped into equivalog families, and otherwise into protein families with other hierarchically defined homology types. This definition is provided at the Institute for Genomic Research (TIGR) website, www.tigr.org; "Terms associated with TIGRFAMs".

The term "variant", as used herein, may refer to polynucleotides or polypeptides that differ from the presently disclosed polynucleotides or polypeptides, respectively, in sequence from each other, and as set forth below.

With regard to polynucleotide variants, differences between presently disclosed polynucleotides and their variants are limited so that the nucleotide sequences of the former and the latter are closely similar overall and, in many regions, identical. The degeneracy of the genetic code dictates that many different variant polynucleotides can encode identical and/or substantially similar polypeptides in addition to those sequences illustrated in the Sequence Listing. Due to this degeneracy, differences between presently disclosed polynucleotides and variant nucleotide sequences may be silent in any given region or over the entire length of the polypeptide (i.e., the amino acids encoded by the polynucleotide

are the same, and the variant polynucleotide sequence thus encodes the same amino acid sequence in that region or entire length of the presently disclosed polynucleotide. Variant nucleotide sequences may encode different amino acid sequences, in which case such nucleotide differences will result in amino acid substitutions, additions, deletions, insertions, truncations or fusions with respect to the similar disclosed polynucleotide sequences. These variations result in polynucleotide variants encoding polypeptides that share at least one functional characteristic (i.e., a presently disclosed transcription factor and a variant will confer at least one of the same functions to a plant).

Within the scope of the invention is a variant of a nucleic acid listed in the Sequence Listing , that is, one having a sequence that differs from the one of the polynucleotide sequences in the Sequence Listing, or a complementary sequence, that encodes a functionally equivalent polypeptide (i.e., a polypeptide having some degree of equivalent or similar biological activity) but differs in sequence from the sequence in the Sequence Listing, due to degeneracy in the genetic code.

"Allelic variant" or "polynucleotide allelic variant" refers to any of two or more alternative forms of a gene occupying the same chromosomal locus. Allelic variation arises naturally through mutation, and may result in phenotypic polymorphism within populations. Gene mutations may be "silent" or may encode polypeptides having altered amino acid sequences. "Allelic variant" and "polypeptide allelic variant" may also be used with respect to polypeptides, and in this case the terms refer to a polypeptide encoded by an allelic variant of a gene.

"Splice variant" or "polynucleotide splice variant" as used herein refers to alternative forms of RNA transcribed from a gene. Splice variation naturally occurs as a result of alternative sites being spliced within a single transcribed RNA molecule or between separately transcribed RNA molecules, and may result in several different forms of mRNA transcribed from the same gene. Thus, splice variants may encode polypeptides having different amino acid sequences, which, in the present context, will have at least one similar function in the organism (splice variation may also give rise to distinct polypeptides having different functions). "Splice variant" or "polypeptide splice variant" may also refer to a polypeptide encoded by a splice variant of a transcribed mRNA.

As used herein, "polynucleotide variants" may also refer to polynucleotide sequences that encode paralogs and orthologs of the presently disclosed polypeptide sequences. "Polypeptide variants" may refer to polypeptide sequences that are paralogs and orthologs of the presently disclosed polypeptide sequences.

Differences between presently disclosed polypeptides and polypeptide variants are limited so that the sequences of the former and the latter are closely similar overall and, in many regions, identical. Presently disclosed polypeptide sequences and similar polypeptide variants may differ in amino acid sequence by one or more substitutions, additions, deletions, fusions and truncations, which may be present in any combination. These differences may produce silent changes and result in a functionally equivalent transcription factor. Thus, it will be readily appreciated by those of skill in the art, that any of a variety of polynucleotide sequences is capable of encoding the transcription factors and transcription

factor homolog polypeptides of the invention. A polypeptide sequence variant may have "conservative" changes, wherein a substituted amino acid has similar structural or chemical properties. Deliberate amino acid substitutions may thus be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the functional or biological activity of the transcription factor is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid, positively charged amino acids may include lysine and arginine, and amino acids with uncharged polar head groups having similar hydrophilicity values may include leucine, isoleucine, and valine; glycine and alanine; asparagine and glutamine; serine and threonine; and phenylalanine and tyrosine. For more detail on conservative substitutions, see Table 5. More rarely, a variant may have "non-conservative" changes, e.g., replacement of a glycine with a tryptophan. Similar minor variations may also include amino acid deletions or insertions, or both. Related polypeptides may comprise, for example, additions and/or deletions of one or more N-linked or O-linked glycosylation sites, or an addition and/or a deletion of one or more cysteine residues. Guidance in determining which and how many amino acid residues may be substituted, inserted or deleted without abolishing functional or biological activity may be found using computer programs well known in the art, for example, DNASTAR software (see USPN 5,840,544).

The term "plant" includes whole plants, shoot vegetative organs/structures (*e.g.*, leaves, stems and tubers), roots, flowers and floral organs/structures (*e.g.*, bracts, sepals, petals, stamens, carpels, anthers and ovules), seed (including embryo, endosperm, and seed coat) and fruit (the mature ovary), plant tissue (*e.g.*, vascular tissue, ground tissue, and the like) and cells (*e.g.*, guard cells, egg cells, and the like), and progeny of same. The class of plants that can be used in the method of the invention is generally as broad as the class of higher and lower plants amenable to transformation techniques, including angiosperms (monocotyledonous and dicotyledonous plants), gymnosperms, ferns, horsetails, psilophytes, lycophytes, bryophytes, and multicellular algae. (See for example, Figure 1, adapted from Daly et al. (2001) *Plant Physiol.* 127: 1328-1333; Figure 2, adapted from Ku et al. (2000) *Proc. Natl. Acad. Sci.* 97: 9121-9126; and see also Tudge, in The Variety of Life, Oxford University Press, New York, NY (2000) pp. 547-606).

A "transgenic plant" refers to a plant that contains genetic material not found in a wild-type plant of the same species, variety or cultivar. The genetic material may include a transgene, an insertional mutagenesis event (such as by transposon or T-DNA insertional mutagenesis), an activation tagging sequence, a mutated sequence, a homologous recombination event or a sequence modified by chimeraplasty. Typically, the foreign genetic material has been introduced into the plant by human manipulation, but any method can be used as one of skill in the art recognizes.

A transgenic plant may contain an expression vector or cassette. The expression cassette typically comprises a polypeptide-encoding sequence operably linked (i.e., under regulatory control of) to appropriate inducible or constitutive regulatory sequences that allow for the expression of polypeptide. The expression cassette can be introduced into a plant by transformation or by breeding after

transformation of a parent plant. A plant refers to a whole plant, including seedlings and mature plants, as well as to a plant part, such as seed, fruit, leaf, or root, plant tissue, plant cells or any other plant material, e.g., a plant explant, as well as to progeny thereof, and to *in vitro* systems that mimic biochemical or cellular components or processes in a cell.

"Fragment", with respect to a polynucleotide, refers to a clone or any part of a polynucleotide molecule that retains a usable, functional characteristic. Useful fragments include oligonucleotides and polynucleotides that may be used in hybridization or amplification technologies or in the regulation of replication, transcription or translation. A polynucleotide fragment" refers to any subsequence of a polynucleotide, typically, of at least about 9 consecutive nucleotides, preferably at least about 30 nucleotides, more preferably at least about 50 nucleotides, of any of the sequences provided herein. Exemplary polynucleotide fragments are the first sixty consecutive nucleotides of the transcription factor polynucleotides listed in the Sequence Listing. Exemplary fragments also include fragments that comprise a region that encodes a conserved domain of a transcription factor.

Fragments may also include subsequences of polypeptides and protein molecules, or a subsequence of the polypeptide. Fragments may have uses in that they may have antigenic potential. In some cases, the fragment or domain is a subsequence of the polypeptide that performs at least one biological function of the intact polypeptide in substantially the same manner, or to a similar extent, as does the intact polypeptide. For example, a polypeptide fragment can comprise a recognizable structural motif or functional domain such as a DNA-binding site or domain that binds to a DNA promoter region, an activation domain, or a domain for protein-protein interactions, and may initiate transcription. Fragments can vary in size from as few as 3 amino acids to the full length of the intact polypeptide, but are preferably at least about 30 amino acids in length and more preferably at least about 60 amino acids in length. Exemplary polypeptide fragments are the first twenty consecutive amino acids of a mammalian protein encoded by are the first twenty consecutive amino acids of the transcription factor polypeptides listed in the Sequence Listing.

Exemplary fragments also include fragments that comprise a conserved domain of a transcription factor. An example of such an exemplary fragment would include amino acid residues 20-110 of G481 (SEQ ID NO: 88), as noted in Table 8.

The invention also encompasses production of DNA sequences that encode transcription factors and transcription factor derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic sequence may be inserted into any of the many available expression vectors and cell systems using reagents well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding transcription factors or any fragment thereof.

A "conserved domain" or "conserved region" as used herein refers to a region in heterologous polynucleotide or polypeptide sequences where there is a relatively high degree of sequence identity between the distinct sequences.

With respect to polynucleotides encoding presently disclosed transcription factors, a conserved region is preferably at least 10 base pairs (bp) in length.

A "conserved domain", with respect to presently disclosed polypeptides refers to a domain within a transcription factor family that exhibits a higher degree of sequence homology, such as at least 26% sequence similarity, at least 16% sequence identity, preferably at least 40% sequence identity, preferably at least 65% sequence identity including conservative substitutions, and more preferably at least 80% sequence identity, and even more preferably at least 85%, or at least about 86%, or at least about 87%, or at least about 88%, or at least about 90%, or at least about 95%, or at least about 98% amino acid residue sequence identity of a polypeptide of consecutive amino acid residues. A fragment or domain can be referred to as outside a conserved domain, outside a consensus sequence, or outside a consensus DNA-binding site that is known to exist or that exists for a particular transcription factor class, family, or sub-family. In this case, the fragment or domain will not include the exact amino acids of a consensus sequence or consensus DNA-binding site of a transcription factor class, family or sub-family, or the exact amino acids of a particular transcription factor consensus sequence or consensus DNA-binding site. Furthermore, a particular fragment, region, or domain of a polypeptide, or a polynucleotide encoding a polypeptide, can be "outside a conserved domain" if all the amino acids of the fragment, region, or domain fall outside of a defined conserved domain(s) for a polypeptide or protein. Sequences having lesser degrees of identity but comparable biological activity are considered to be equivalents.

As one of ordinary skill in the art recognizes, conserved domains of transcription factors may be identified as regions or domains of identity to a specific consensus sequence (see, for example, Riechmann et al. (2000) *supra*). Thus, by using alignment methods well known in the art, the conserved domains of the plant transcription factors for each of the following may be determined: the AP2 (APETALA2) domain transcription factor family (Riechmann and Meyerowitz (1998) *supra*; the MYB transcription factor family (ENBib; Martin and Paz-Ares (1997) *supra*); the MADS domain transcription factor family (Riechmann and Meyerowitz (1997) *supra*); the WRKY protein family (Ishiguro and Nakamura (1994) *supra*); the ankyrin-repeat protein family (Zhang et al. (1992) *supra*); the zinc finger protein (Z) family (Klug and Schwabe (1995) *supra*; Takatsuji (1998) *supra*); the homeobox (HB) protein family (Buerglin (1994) *supra*); the CAAT-element binding proteins (Forsburg and Guarente (1989) *supra*); the squamosa promoter binding proteins (SPB) (Klein et al. (1996) *supra*); the NAM protein family (Souer et al. (1996) *supra*); the IAA/AUX proteins (Abel et al. (1995) *supra*); the HLH/MYC protein family (Littlewood et al. (1994) *supra*); the DNA-binding protein (DBP) family (Tucker et al. (1994) *supra*); the bZIP family of transcription factors (Foster et al. (1994) *supra*); the Box P-binding protein (the BPF-1) family (da Costa e Silva et al. (1993) *supra*); the high mobility group (HMG) family (Bustin and Reeves (1996) *supra*); the scarecrow (SCR) family (Di Laurenzio et al. (1996) *supra*); the GF14 family (Wu et al. (1997) *supra*); the polycomb (PCOMB) family (Goodrich et al. (1997) *supra*); the teosinte branched (TEO) family (Luo et al. (1996) *supra*); the ABI3 family (Giraudat et al. (1992) *supra*); the triple helix (TH) family (Dehesh et al. (1990) *supra*); the EIL family

(Chao et al. (1997) *Cell supra*); the AT-HOOK family (Reeves and Nissen (1990 *supra*); the S1FA family (Zhou et al. (1995) *supra*); the bZIPT2 family (Lu and Ferl (1995) *supra*); the YABBY family (Bowman et al. (1999) *supra*); the PAZ family (Bohmert et al. (1998) *supra*); a family of miscellaneous (MISC) transcription factors including the DPBF family (Kim et al. (1997) *supra*) and the SPF1 family (Ishiguro and Nakamura (1994) *supra*); the GARP family (Hall et al. (1998) *supra*), the TUBBY family (Boggin et al. (1999) *supra*), the heat shock family (Wu (1995 *supra*), the ENBP family (Christiansen et al. (1996) *supra*), the RING-zinc family (Jensen et al. (1998) *supra*), the PDBP family (Janik et al. (1989) *supra*), the PCF family (Cubas et al. (1999) *supra*), the SRS (SHI-related) family (Fridborg et al. (1999) *supra*), the CPP (cysteine-rich polycomb-like) family (Cvitanich et al. (2000) *supra*), the ARF (auxin response factor) family (Ulmasov et al. (1999) *supra*), the SWI/SNF family (Collingwood et al. (1999) *supra*), the ACBF family (Seguin et al. (1997) *supra*), PCGL (CG-1 like) family (da Costa e Silva et al. (1994) *supra*) the ARID family (Vazquez et al. (1999) *supra*), the Jumonji family, (Balciunas et al. (2000) *supra*), the bZIP-NIN family (Schauser et al. (1999) *supra*), the E2F family Kaelin et al. (1992) *supra*) and the GRF-like family (Knaap et al (2000) *supra*).

The conserved domains for each of polypeptides of SEQ ID NO: 2N, wherein N = 1- 229, are listed in Table 8 as described in Example VII. Also, many of the polypeptides of Table 8 have conserved domains specifically indicated by start and stop sites. A comparison of the regions of the polypeptides in SEQ ID NO: 2N, wherein N = 1- 229, or of those in Table 8, allows one of skill in the art to identify conserved domain(s) for any of the polypeptides listed or referred to in this disclosure, including those in Tables 7 - 11.

As used herein, a "gene" is a functional unit of inheritance, and in physical terms is a particular segment or sequence of nucleotides along a molecule of DNA (or RNA, in the case of RNA viruses) involved in producing a functional RNA molecule, such as one used for a structural or regulatory role, or a polypeptide chain, such as one used for a structural or regulatory role (an example of the latter would be transcription regulation, as by a transcription factor polypeptide). Polypeptides may then be subjected to subsequent processing such as splicing and/or folding to obtain a functional polypeptide. A gene may be isolated, partially isolated, or be found with an organism's genome. By way of example, a transcription factor gene encodes a transcription factor polypeptide, which may be functional with or without additional processing to function as an initiator of transcription.

Operationally, genes may be defined by the cis-trans test, a genetic test that determines whether two mutations occur in the same gene and which may be used to determine the limits of the genetically active unit (Rieger et al. (1976) Glossary of Genetics and Cytogenetics: Classical and Molecular, 4th ed., Springer Verlag. Berlin). A gene generally includes regions preceding ("leaders"; upstream) and following ("trailers"; downstream) of the coding region. A gene may also include intervening, non-coded sequences, referred to as "introns", which are located between individual coding segments, referred to as "exons". Most genes have an identifiable associated promoter region, a regulatory sequence

5' or upstream of the transcription initiation codon. The function of a gene may also be regulated by enhancers, operators, and other regulatory elements.

A "trait" refers to a physiological, morphological, biochemical, or physical characteristic of a plant or particular plant material or cell. In some instances, this characteristic is visible to the human eye, such as seed or plant size, or can be measured by biochemical techniques, such as detecting the protein, starch, or oil content of seed or leaves, or by observation of a metabolic or physiological process, e.g. by measuring uptake of carbon dioxide, or by the observation of the expression level of a gene or genes, e.g., by employing Northern analysis, RT-PCR, microarray gene expression assays, or reporter gene expression systems, or by agricultural observations such as stress tolerance, yield, or pathogen tolerance. Any technique can be used to measure the amount of, comparative level of, or difference in any selected chemical compound or macromolecule in the transgenic plants, however.

"Trait modification" refers to a detectable difference in a characteristic in a plant ectopically expressing a polynucleotide or polypeptide of the present invention relative to a plant not doing so, such as a wild-type plant. In some cases, the trait modification can be evaluated quantitatively. For example, the trait modification can entail at least about a 2% increase or decrease in an observed trait (difference), at least a 5% difference, at least about a 10% difference, at least about a 20% difference, at least about a 30%, at least about a 50%, at least about a 70%, or at least about a 100%, or an even greater difference compared with a wild-type plant. It is known that there can be a natural variation in the modified trait. Therefore, the trait modification observed entails a change of the normal distribution of the trait in the plants compared with the distribution observed in wild-type plant.

The term "transcript profile" refers to the expression levels of a set of genes in a cell in a particular state, particularly by comparison with the expression levels of that same set of genes in a cell of the same type in a reference state. For example, the transcript profile of a particular transcription factor in a suspension cell is the expression levels of a set of genes in a cell overexpressing that transcription factor compared with the expression levels of that same set of genes in a suspension cell that has normal levels of that transcription factor. The transcript profile can be presented as a list of those genes whose expression level is significantly different between the two treatments, and the difference ratios. Differences and similarities between expression levels may also be evaluated and calculated using statistical and clustering methods.

"Wild type", as used herein, refers to a cell, tissue or plant that has not been genetically modified to knock out or overexpress one or more of the presently disclosed transcription factors. Wild-type cells, tissue or plants may be used as controls to compare levels of expression and the extent and nature of trait modification with modified (e.g., transgenic) cells, tissue or plants in which transcription factor expression is altered or ectopically expressed by, for example, knocking out or overexpressing a gene.

"Ectopic expression" or "altered expression" in reference to a polynucleotide indicates that the pattern of expression in, e.g., a transgenic plant or plant tissue, is different from the expression pattern in a wild-type plant or a reference plant of the same species. The pattern of expression may also be

compared with a reference expression pattern in a wild-type plant of the same species. For example, the polynucleotide or polypeptide is expressed in a cell or tissue type other than a cell or tissue type in which the sequence is expressed in the wild-type plant, or by expression at a time other than at the time the sequence is expressed in the wild-type plant, or by a response to different inducible agents, such as hormones or environmental signals, or at different expression levels (either higher or lower) compared with those found in a wild-type plant. Altered expression may be achieved by, for example, transformation of a plant with an expression cassette having a constitutive or inducible promoter element associated with a transcription factor gene. The resulting expression pattern can thus constitutive or inducible, and be stable or transient. Altered or ectopic expression may also refer to altered expression patterns that are produced by lowering the levels of expression to below the detection level or completely abolishing expression by, for example, knocking out a gene's expression by disrupting expression or regulation of the gene with an insertion element.

In reference to a polypeptide, the term "ectopic expression or altered expression" further may relate to altered activity levels resulting from the interactions of the polypeptides with exogenous or endogenous modulators or from interactions with factors or as a result of the chemical modification of the polypeptides.

The term "overexpression" as used herein refers to a greater expression level of a gene in a plant, plant cell or plant tissue, compared to expression in a wild-type plant, cell or tissue, at any developmental or temporal stage for the gene. Overexpression can occur when, for example, the genes encoding one or more transcription factors are under the control of a strong expression signal, such as one of the promoters described herein (e.g., the cauliflower mosaic virus 35S transcription initiation region). Overexpression may occur throughout a plant or in specific tissues of the plant, depending on the promoter used, as described below.

Overexpression may take place in plant cells normally lacking expression of polypeptides functionally equivalent or identical to the present transcription factors. Overexpression may also occur in plant cells where endogenous expression of the present transcription factors or functionally equivalent molecules normally occurs, but such normal expression is at a lower level than in the organism or tissues of the overexpressor. Overexpression thus results in a greater than normal production, or "overproduction" of the transcription factor in the plant, cell or tissue.

The term "phase change" refers to a plant's progression from embryo to adult, and, by some definitions, the transition wherein flowering plants gain reproductive competency. It is believed that phase change occurs either after a certain number of cell divisions in the shoot apex of a developing plant, or when the shoot apex achieves a particular distance from the roots. Thus, altering the timing of phase changes may affect a plant's size, which, in turn, may affect yield and biomass.

Traits That May Be Modified in Overexpressing or Knock-out Plants

Trait modifications of particular interest include those to seed (such as embryo or endosperm), fruit, root, flower, leaf, stem, shoot, seedling or the like, including: enhanced tolerance to environmental conditions including freezing, chilling, heat, drought, water saturation, radiation and ozone; improved tolerance to microbial, fungal or viral diseases; improved tolerance to pest infestations, including insects, nematodes, mollicutes, parasitic higher plants or the like; decreased herbicide sensitivity; improved tolerance of heavy metals or enhanced ability to take up heavy metals; improved growth under poor photoconditions (e.g., low light and/or short day length), or changes in expression levels of genes of interest. Other phenotype that can be modified relate to the production of plant metabolites, such as variations in the production of taxol, tocopherol, tocotrienol, sterols, phytosterols, vitamins, wax monomers, anti-oxidants, amino acids, lignins, cellulose, tannins, prenyllipids (such as chlorophylls and carotenoids), glucosinolates, and terpenoids, enhanced or compositionally altered protein or oil production (especially in seeds), or modified sugar (insoluble or soluble) and/or starch composition. Physical plant characteristics that can be modified include cell development (such as the number of trichomes), fruit and seed size and number, yields of plant parts such as stems, leaves, inflorescences, and roots, the stability of the seeds during storage, characteristics of the seed pod (e.g., susceptibility to shattering), root hair length and quantity, internode distances, or the quality of seed coat. Plant growth characteristics that can be modified include growth rate, germination rate of seeds, vigor of plants and seedlings, leaf and flower senescence, male sterility, apomixis, flowering time, flower abscission, rate of nitrogen uptake, osmotic sensitivity to soluble sugar concentrations, biomass or transpiration characteristics, as well as plant architecture characteristics such as apical dominance, branching patterns, number of organs, organ identity, organ shape or size.

Transcription Factors Modify Expression of Endogenous Genes

Expression of genes that encode transcription factors that modify expression of endogenous genes, polynucleotides, and proteins are well known in the art. In addition, transgenic plants comprising isolated polynucleotides encoding transcription factors may also modify expression of endogenous genes, polynucleotides, and proteins. Examples include Peng et al. (1997) *Genes and Development* 11: 3194-3205, and Peng et al. (1999) *Nature* 400: 256-261. In addition, many others have demonstrated that an *Arabidopsis* transcription factor expressed in an exogenous plant species elicits the same or very similar phenotypic response. See, for example, Fu et al. (2001) *Plant Cell* 13: 1791-1802; Nandi et al. (2000, *Curr. Biol.* 10: 215-218; Coupland (1995) *Nature* 377: 482-483; and Weigel and Nilsson (1995) *Nature* 377: 482-500.

In another example, Mandel et al. (1992) *Cell* 71-133-143 and Suzuki et al. (2001) *Plant J.* 28: 409-418, teach that a transcription factor expressed in another plant species elicits the same or very similar phenotypic response of the endogenous sequence, as often predicted in earlier studies of

Arabidopsis transcription factors in *Arabidopsis* (see Mandel et al. (1992) *supra*; Suzuki et al. (2001) *supra*).

Other examples include Müller et al. (2001) *Plant J.* 28: 169-179; Kim et al. (2001) *Plant J.* 25: 247-259; Kyozuka and Shimamoto (2002) *Plant Cell Physiol.* 43: 130-135; Boss and Thomas (2002) *Nature* 416: 847-850; He et al. (2000) *Transgenic Res.* 9: 223-227; and Robson et al. (2001) *Plant J.* 28: 619-631.

In yet another example, Gilmour et al. (1998) *Plant J.* 16: 433-442, teach an *Arabidopsis* AP2 transcription factor, CBF1 (SEQ ID NO: 1956), which, when overexpressed in transgenic plants, increases plant freezing tolerance. Jaglo et al. (2001) *Plant Physiol.* 127: 910-917, further identified sequences in *Brassica napus* which encode CBF-like genes and that transcripts for these genes accumulated rapidly in response to low temperature. Transcripts encoding CBF-like proteins were also found to accumulate rapidly in response to low temperature in wheat, as well as in tomato. An alignment of the CBF proteins from *Arabidopsis*, *B. napus*, wheat, rye, and tomato revealed the presence of conserved consecutive amino acid residues, PKK/RPAGRxKFxETRHP and DSAWR, that bracket the AP2/EREBP DNA binding domains of the proteins and distinguish them from other members of the AP2/EREBP protein family. (See Jaglo et al. *supra*).

Gao et al. (2002) *Plant Molec. Biol.* 49: 459-471) have recently described four CBF transcription factors from *Brassica napus*: BNCBFs 5, 7, 16 and 17. They note that the first three CBFs (GenBank Accession Numbers AAM18958, AAM18959, and AAM18960, respectively) are very similar to *Arabidopsis* CBF1, whereas BNCBF17 (GenBank Accession Number AAM18961) is similar but contains two extra regions of 16 and 21 amino acids in its acidic activation domain. All four *B. napus* CBFs accumulate in leaves of the plants after cold-treatment, and BNCBFs 5, 7, 16 accumulated after salt stress treatment. The authors concluded that these BNCBFs likely function in low-temperature responses in *B. napus*.

In a functional study of CBF genes,, Hsieh et al. ((2002) *Plant Physiol.* 129: 1086-1094) found that heterologous expression of *Arabidopsis* CBF1 in tomato plants confers increased tolerance to chilling and considerable tolerance to oxidative stress, which suggested to the authors that ectopic *Arabidopsis* CBF1 expression may induce several tomato stress responsive genes to protect the plants.

Transcription factors mediate cellular responses and control traits through altered expression of genes containing cis-acting nucleotide sequences that are targets of the introduced transcription factor. It is well appreciated in the Art that the effect of a transcription factor on cellular responses or a cellular trait is determined by the particular genes whose expression is either directly or indirectly (e.g., by a cascade of transcription factor binding events and transcriptional changes) altered by transcription factor binding. In a global analysis of transcription comparing a standard condition with one in which a transcription factor is overexpressed, the resulting transcript profile associated with transcription factor overexpression is related to the trait or cellular process controlled by that transcription factor. For example, the PAP2 gene (and other genes in the MYB family) have been shown to control anthocyanin

biosynthesis through regulation of the expression of genes known to be involved in the anthocyanin biosynthetic pathway (Bruce et al. (2000) *Plant Cell* 12: 65-79; and Borevitz et al. (2000) *Plant Cell* 12: 2383-2393). Further, global transcript profiles have been used successfully as diagnostic tools for specific cellular states (e.g., cancerous vs. non-cancerous; Bhattacharjee et al. (2001) *Proc. Natl. Acad. Sci.* USA 98: 13790-13795; and Xu et al. (2001) *Proc Natl Acad Sci*, USA 98: 15089-15094). Consequently, it is evident to one skilled in the art that similarity of transcript profile upon overexpression of different transcription factors would indicate similarity of transcription factor function.

Polypeptides and Polynucleotides of the Invention

The present invention provides, among other things, transcription factors (TFs), and transcription factor homolog polypeptides, and isolated or recombinant polynucleotides encoding the polypeptides, or novel sequence variant polypeptides or polynucleotides encoding novel variants of transcription factors derived from the specific sequences provided here. These polypeptides and polynucleotides may be employed to modify a plant's characteristics.

Exemplary polynucleotides encoding the polypeptides of the invention were identified in the *Arabidopsis thaliana* GenBank database using publicly available sequence analysis programs and parameters. Sequences initially identified were then further characterized to identify sequences comprising specified sequence strings corresponding to sequence motifs present in families of known transcription factors. In addition, further exemplary polynucleotides encoding the polypeptides of the invention were identified in the plant GenBank database using publicly available sequence analysis programs and parameters. Sequences initially identified were then further characterized to identify sequences comprising specified sequence strings corresponding to sequence motifs present in families of known transcription factors. Polynucleotide sequences meeting such criteria were confirmed as transcription factors.

Additional polynucleotides of the invention were identified by screening *Arabidopsis thaliana* and/or other plant cDNA libraries with probes corresponding to known transcription factors under low stringency hybridization conditions. Additional sequences, including full length coding sequences were subsequently recovered by the rapid amplification of cDNA ends (RACE) procedure, using a commercially available kit according to the manufacturer's instructions. Where necessary, multiple rounds of RACE are performed to isolate 5' and 3' ends. The full-length cDNA was then recovered by a routine end-to-end polymerase chain reaction (PCR) using primers specific to the isolated 5' and 3' ends. Exemplary sequences are provided in the Sequence Listing.

The polynucleotides of the invention can be or were ectopically expressed in overexpressor or knockout plants and the changes in the characteristic(s) or trait(s) of the plants observed. Therefore, the polynucleotides and polypeptides can be employed to improve the characteristics of plants.

The polynucleotides of the invention can be or were ectopically expressed in overexpressor plant cells and the changes in the expression levels of a number of genes, polynucleotides, and/or proteins of

the plant cells observed. Therefore, the polynucleotides and polypeptides can be employed to change expression levels of a genes, polynucleotides, and/or proteins of plants.

Specific examples of the polypeptides and polynucleotides of the invention and experimental observations made after modifying their expression in plants may be found in the following text and tables.

CCAAT-element binding protein transcription factor family

G481 (AT2G38880) and G482 are members of the HAP3 sub-group of the CCAAT binding factor family (CAAT). G481, G481 and their related sequences were included a program to test the ability of these sequences to confer the same drought-related abiotic stress previously observed by us in 35S::G481 lines.

The CAAT family of transcription factors, also be referred to as the "CCAAT" or "CCAAT-box" family, are characterized by their ability to bind to the CCAAT-box element located 80 to 300 bp 5' from a transcription start site (Gelinas et al. (1985) *Nature* 313: 323-325). The CCAAT-box is a conserved cis-acting regulatory element with the consensus sequence CCAAT that is found in the promoters of genes from all eukaryotic species. The element can act in either orientation, alone or as multimeric regions with possible cooperation with other cis regulatory elements (Tasanen et al. (1992) (*J. Biol. Chem.* 267: 11513-11519). It has been estimated that 25% of eukaryotic promoters harbor this element (Bucher (1988) *J. Biomol. Struct. Dyn.* 5: 1231-1236). CCAAT-box elements have been shown to function in the regulation of gene expression in plants (Rieping and Schoffl (1992) *Mol. Gen. Genet.* 231: 226-232; Kehoe et al. (1994) *Plant Cell* 6: 1123-1134; Ito et al. (1995) *Plant Cell Physiol.* 36: 1281-1289). Several reports have described the importance of the CCAAT-binding element for regulated expression; including the regulation of genes that are responsive to light (Kusnetsov et al. (1999) *J. Biol. Chem.* 274: 36009-36014; Carre and Kay (1995) *Plant Cell* 7: 2039-2051) as well as stress (Rieping and Schoffl (1992) *supra*). Specifically, a CCAAT-box motif was shown to be important for the light regulated expression of the CAB2 promoter in *Arabidopsis*, however, the proteins that bind to the site were not identified (Carre and Kay (1995) *supra*). To date, no specific *Arabidopsis* CCAAT-box binding protein has been functionally associated with its corresponding target genes. In October of 2002 at an EPSO meeting on Plant Networks, a seminar was given by Detlef Weigel (Tuebingen) on the control of the AGAMOUS (a floral organ identity gene) gene in *Arabidopsis*. In order to find important cis-elements that regulate AGAMOUS activity, he aligned the promoter regions from 29 different Brassicaceae species and showed that there were two highly conserved regions; one well characterized site that binds LEAFY/WUS heterodimers and another putative CCAAT-box binding motif. We have discovered several CCAAT-box genes that regulate flowering time and are candidates for binding to the AGAMOUS promoter. One of these genes, G485, is a HAP3-like protein that is closely related to G481. Gain of function and loss of function studies on G485 reveal opposing effects on flowering time,

indicating that the gene is both sufficient to act as a floral activator, and is also necessary in that role within the plant.

The first proteins identified that bind to the CCAAT-box element were identified in yeast. The CCAAT-box transcription factors bind as hetero-tetrameric complex called the HAP complex (heme activator protein complex) or the CCAAT binding factor (Forsburg and Guarente (1988) *Mol. Cell Biol.* 8: 647-654). The HAP complex in yeast is composed of at least four subunits, HAP2, HAP3, HAP4 and HAP5. In addition, the proteins that make up the HAP2,3,4,5 complex are represented by single genes. Their function is specific for the activation of genes involved in mitochondrial biogenesis and energy metabolism (Dang et al. (1996) *Mol. Microbiol.* 22:681-692). In mammals, the CCAAT binding factor is a trimeric complex consisting of NF-YA (HAP2-like), NF-YB (HAP3-like) and NF-YC (HAP5-like) subunits (Maity and de Crombrugghe (1998) *Trends Biochem. Sci.* 23: 174-178). In plants, analogous members of the CCAAT binding factor complex are represented by small gene families, and it is likely that these genes play a more complex role in regulating gene transcription. In *Arabidopsis* there are ten members of the HAP2 subfamily, ten members of the HAP3 subfamily, thirteen members of the HAP5 subfamily. Plants and mammals, however, do not appear to have a protein equivalent of HAP4 of yeast. HAP4 is not required for DNA binding in yeast although it provides the primary activation domain for the complex (McNabb et al. (1995) *Genes Dev.* 9: 47-58; Olesen and Guarente (1990) *Genes Dev.* 4, 1714-1729).

In mammals, the CCAAT-box element is found in the promoters of many genes and it is therefore been proposed that CCAAT binding factors serve as general transcriptional regulators that influence the frequency of transcriptional initiation (Maity and de Crombrugghe (1998) *supra*). CCAAT binding factors, however, can serve to regulate target promoters in response to environmental cues and it has been demonstrated that assembly of CCAAT binding factors on target promoters occurs in response to a variety signals (Myers et al. (1986) Myers et al. (1986) *Science* 232: 613-618; Maity and de Crombrugghe (1998) *supra*; Bezhani et al. (2001) *J. Biol. Chem.* 276: 23785-23789). Mammalian CP1 and NF-Y are both heterotrimeric CCAAT binding factor complexes (Johnson and McKnight (1989) *Ann. Rev. Biochem.* 58: 799-839. Plant CCAAT binding factors are assumed to be trimeric, as is the case in mammals, however, they could associate with other transcription factors on target promoters as part of a larger complex. The CCAAT box is generally found in close proximity of other promoter elements and it is generally accepted that the CCAAT binding factor functions synergistically with other transcription factors in the regulation of transcription. In addition, it has recently been shown that a HAP3-like protein from rice, OsNF-YB1, interacts with a MADS-box protein OsMADS18 in vitro (Masiero et al. (2002) *J. Biol. Chem.* 277: 26429-26435). It was also shown that the in vitro ternary complex between these two types of transcription factors requires that both; OsNF-YB1 form a dimer with a HAP5-like protein, and that OsMADS18 form a heterodimer with another MADS-box protein. Interestingly, the OsNF-YB1/HAP5 protein dimer is incapable of interacting with HAP2-like subunits and therefore cannot bind the CCAAT element. The authors therefore speculate that there is a select set of HAP3-like proteins in

plants that act on non-CCAAT promoter elements by virtue of their interaction with other non-CCAAT transcription factors (Masiero et al. (2002) *supra*). In support of this, HAP3/HAP5 subunit dimers have been shown to be able to interact with TFIID in the absence of HAP2 subunits (Romier et al. (2003) *J. Biol. Chem.* 278: 1336-1345).

The CCAAT-box motif is found in the promoters of a variety of plant genes. In addition, the expression pattern of many of the HAP-like genes in *Arabidopsis* shows developmental regulation. We have used RT-PCR to analyze the endogenous expression of 31 of the 34 CCAAT-box proteins . Our findings suggest that while most of the CCAAT-box gene transcripts are found ubiquitously throughout the plant, in more than half of the cases, the genes are predominantly expressed in flower, embryo and/or silique tissues. Cell-type specific localization of the CCAAT genes in *Arabidopsis* would be very informative and could help determine the activity of various CCAAT genes in the plant.

Genetic analysis has determined the function of one *Arabidopsis* CCAAT gene, LEAFY COTYLEDON (LEC1). LEC1 is a HAP3 subunit gene that accumulates only during seed development. *Arabidopsis* plants carrying a mutation in the LEC1 gene display embryos that are intolerant to desiccation and that show defects in seed maturation (Lotan et al. (1998) *Cell* 93: 1195-1205). This phenotype can be rescued if the embryos are allowed to grow before the desiccation process occurs during normal seed maturation. This result suggests LEC1 has a role in allowing the embryo to survive desiccation during seed maturation. The mutant plants also possess trichomes, or epidermal hairs on their cotyledons, a characteristic that is normally restricted to adult tissues like leaves and stems. Such an effect suggests that LEC1 also plays a role in specifying embryonic organ identity. In addition to the mutant analysis, the ectopic expression (unregulated overexpression) of the wild type LEC1 gene induces embryonic programs and embryo development in vegetative cells consistent with its role in coordinating higher plant embryo development. The ortholog of LEC1 has been identified recently in maize. The expression pattern of ZmLEC1 in maize during somatic embryo development is similar to that of LEC1 in *Arabidopsis* during zygotic embryo development (Zhang et al. (2002) *Planta* 215:191-194).

Matching the CCAAT transcription factors with target promoters and the analysis of the knockout and overexpression mutant phenotypes will help sort out whether these proteins act specifically or non-specifically in the control of plant pathways. The fact that CCAAT-box elements are not present in most plant promoters suggests that plant CCAAT binding factors most likely do not function as general components of the transcriptional machinery. In addition, the very specific role of the LEC1 protein in plant developmental processes supports the idea that CCAAT-box binding complexes play very specific roles in plant growth and development.

The domain structure of CCAAT-element binding transcription factors and novel conserved domains in *Arabidopsis* and other species

Plant CCAAT binding factors potentially bind DNA as heterotrimers composed of HAP2-like, HAP3-like and HAP5-like subunits. All subunits contain regions that are required for DNA binding and

subunit association. The subunit proteins appear to lack activation domains; therefore, that function must come from proteins with which they interact on target promoters. No proteins that provide the activation domain function for CCAAT binding factors have been identified in plants. In yeast, however, the HAP4 protein provides the primary activation domain (McNabb et al. (1995) *Genes Dev.* 9: 47-58; Olesen and Guarente (1990) *Genes Dev.* 4, 1714-1729).

HAP2-, HAP3- and HAP5-like proteins have two highly conserved sub-domains, one that functions in subunit interaction and the other that acts in a direct association with DNA. Outside these two regions, non-paralogous *Arabidopsis* HAP-like proteins are quite divergent in sequence and in overall length.

The general domain structure of HAP3 proteins is found in Figure 9. HAP3 proteins contain an amino-terminal A domain, a central B domain and a carboxy-terminal C domain. There is very little sequence similarity between HAP3 proteins in the A and C domains; it is therefore reasonable to assume that the A and C domains could provide a degree of functional specificity to each member of the HAP3 subfamily. The B domain is the conserved region that specifies DNA binding and subunit association.

In Figures 10A-10F, HAP3 proteins from *Arabidopsis*, soybean, rice and corn are aligned with G481, with the A, B and C domains and the DNA binding and subunit interaction domains indicated. As can be seen in Figure 10B-10C, the B domain of the non-LEC1-like clade (identified in Figures 6 and 7) may be distinguished by the amino acid residues:

Ser/Gly-Arg-Ile/Leu-Met-Lys-(Xaa)$_2$-Lys/Ile/Val-Pro-Xaa-**Asn**-Ala/Gly-Lys-Ile/Val-Ser/Ala/Gly-**Lys**-Asp/Glu-Ala/Ser-Lys-Glu/Asp/Gln-Thr/Ile-Xaa-Gln-Glu-Cys-Val/Ala-Ser/Thr-Glu-Phe-Ile-Ser-Phe-Ile/Val/His-Thr/Ser-[Pro]-Gly/Ser/Cys-Glu-Ala/Leu-Ser/Ala-Asp/Glu/Gly-Lys/Glu-Cys-Gln/His-Arg/Lys-Glu-Lys/Asn-Arg-Lys-Thr-Ile/Val-**Asn-Gly**-Asp/Glu-Asp-Leu/Ile-Xaa-Trp/Phe-Ala-Met/Ile/Leu-Xaa-Thr/Asn-Leu-Gly-Phe/Leu-Glu/Asp-Xaa-Tyr-(Xaa)$_2$-Pro/Gln/Ala-Leu/Val-Lys/Gly;

where Xaa can be any amino acid. The proline residue that appears in brackets is an additional residue that was found in only one sequence (not shown in Figure 10B). The boldfaced residues that appear here and in the consensus sequences of Figures 10B - 10C in their present positions are uniquely found in the non-LEC1-like clade, and may be used to identify members of this clade. The G482-like subclade may be delineated by the underlined serine residue in its present position here and in the consensus sequence of Figures 10B-10C. More generally, the non-LEC1-like clade is distinguished by a B domain comprising:

Asn-(Xaa)$_4$-Lys-(Xaa)$_{33-34}$-Asn-Gly;

and the G482 subclade is distinguished by a B-domain comprising:

Ser-$(Xaa)_9$-Asn-$(Xaa)_4$-Lys-$(Xaa)_{33-34}$-Asn-Gly.

Overexpression of these polypeptides confers increased abiotic stress tolerance in a transgenic plant, as compared to a non-transformed plant that does not overexpress the polypeptide.

The CCAAT family members under study

G481, G482 and G485 (polynucleotide SEQ ID NOs: 87, 89 and 2009) were chosen for study based on observations that *Arabidopsis* plants overexpressing these genes had resistance to abiotic stresses, such as osmotic stress, and including drought-related stress (see Example XIII, below). G481, G482 and G485 are members of the CCAAT family, proteins that act in a multi-subunit complex and are believed to bind CCAAT boxes in promoters of target genes as trimers or tetramers.

In *Arabidopsis*, three types of CCAAT binding proteins exist: HAP2, HAP3 and HAP5. The G481, G482 and G485 polypeptides, as well as a number of other proteins in the *Arabidopsis* proteome, belong to the HAP3 class. As reported in the scientific literature thus far, only two genes from the HAP3 class have been functionally analyzed to a substantial degree. These are LEAFY COTYLEDON1 (LEC1) and its most closely related subunit, LEC1-LIKE (L1L). LEC1 and L1L are expressed primarily during seed development. Both appear to be essential for embryo survival of desiccation during seed maturation (Kwong et al. (2003) *Plant Cell* 15: 5-18). LEC1 is a critical regulator required for normal development during the early and late phases of embryogenesis that is sufficient to induce embryonic development in vegetative cells. Kwong et al. showed that ten *Arabidopsis* HAP3 subunits can be divided into two classes based on sequence identity in their central, conserved B domain. LEC1 and L1L constitute LEC1-type HAP3 subunits, whereas the remaining HAP3 subunits were designated non-LEC1-type.

Phylogenetic trees based on sequential relatedness of the HAP3 genes are shown in Figure 6 and 7. As can be seen in these figures showing the L1L-related CCAAT transcription factor family, G1364 and G2345 are closely related to G481, and G482 and G485 are more related to G481 than either LEC1 or L1L, the latter two sequences being found on somewhat more distant nodes. The present invention encompasses the G482 subclade of these non-LEC1-like clade of proteins of the L1L-related CCAAT transcription factor family, for which a representative number of monocot and dicot species, including members from dicot and monocot species (for example, *Arabidopsis* sequences G481, G482, G485, G1364 and G2345, soy sequences G3472, G3475, G3476, rice sequences G3395, G3397, G3398, and corn sequences G3434, and G3436), have been shown to confer improved abiotic stress tolerance in plants when overexpressed.

Table 1 shows the polypeptides identified by SEQ ID NO; Gene ID (GID) No.; the transcription factor family to which the polypeptide belongs, and conserved B domains of the polypeptide. The first column shows the polypeptide SEQ ID NO; the second column the species (abbreviated) and identifier (GID or "Gene IDentifier); the third column shows the B domain; the fourth column shows the amino acid coordinates of the conserved domain which were used to determine percentage identity to G481; and the fifth column shows the percentage identity to G481. The sequences are arranged in descending order of percentage identity to G481.

For Tables 1 and 2, homology was determined after aligning the sequences using the methods of Smith and Waterman (1981) *Adv. Appl. Math.* 2: 482-489. After alignment, sequence comparisons between the polypeptides were performed by comparison over a comparison window to identify and compare local regions of sequence similarity. A description of the method is provided in Ausubel et al. (eds.) *Current Protocols in Molecular Biology*, John Wiley & Sons (1998, and supplements through 2001), Altschul et al.(1990). *J. Mol. Biol.* 215: 403-410; and Gish and States (1993) *Nature Genetics* 3: 266-72. The percentage identity reported in these tables is based on the comparison within these windows.

Table 1. Gene families and B domains

Polypeptide SEQ ID NO:	Species/ GID No., Accession No., or Identifier	B Domain	Amino Acid Coordinates for Percent Identity Determination	% ID to CCAAT-box binding conserved domain of G481 (SEQ ID NO: 88)
88	At/G481	REQDRYLPIANISRIMKKALPPNGKI GKDAKDTVQECVSEFISFITSEASDK CQKEKRKTVNGDDLLWAMATLGF EDYLEPLKIYLARYRE	20-110	100%
806	Zm/G3434	REQDRFLPIANISRIMKKAVPANGKI AKDAKETLQECVSEFISFVTSEASD KCQKEKRKTINGDDLLWAMATLGF EEYVEPLKIYLQKYKE	18-108	85%
2102	At/G1364	REQDRFLPIANISRIMKRGLPANGKI AKDAKEIVQECVSEFISFVTSEASDK CQREKRKTINGDDLLWAMATLGFE DYMEPLKVYLMRYRE	29-119	85%
800	Gm/G3475	REQDRFLPIANVSRIMKKALPANAK ISKDAKETVQECVSEFISFITGEASD KCQREKRKTINGDDLLWAMTTLGF EDYVEPLKGYLQRFRE	23-113	84%
2010	At/G485	REQDRFLPIANVSRIMKKALPANAK ISKDAKETVQECVSEFISFITGEASD KCQREKRKTINGDDLLWAMTTLGF EDYVEPLKVYLQKYRE	20-110	84%
798	Gm/G3476	REQDRFLPIANVSRIMKKALPANAK ISKDAKETVQECVSEFISFITGEASD	26-116	84%

		KCQREKRKTINGDDLLWAMTTLGF EEYVEPLKIYLQRFRE		
2172	At/G2345	REQDRFLPIANISRIMKRGLPLNGKI AKDAKETMQECVSEFISFVTSEASD KCQREKRKTINGDDLLWAMATLGF EDYIDPLKVYLMRYRE	28-118	84%
90	At/G482	REQDRFLPIANVSRIMKKALPANAK ISKDAKETMQECVSEFISFVTGEASD KCQKEKRKTINGDDLLWAMTTLGF EDYVEPLKVYLQRFRE	26-116	83%
801	Gm/G3472	REQDRFLPIANVSRIMKKALPANAK ISKEAKETVQECVSEFISFITGEASD KCQKEKRKTINGDDLLWAMTTLGF EEYVEPLKVYLHKYRE	25-115	83%
805	Zm/G3436	REQDRFLPIANVSRIMKKALPANAK ISKDAKETVQECVSEFISFITGEASD KCQREKRKTINGDDLLWAMTTLGF EDYVEPLKLYLHKFRE	20-110	83%
794	Os/G3397	REQDRFLPIANVSRIMKKALPANAK ISKDAKETVQECVSEFISFITGEASD KCQREKRKTINGDDLLWAMTTLGF EDYVDPLKHYLHKFRE	23-113	82%
790	Os/G3395	REQDRFLPIANISRIMKKAVPANGKI AKDAKETLQECVSEFISFVTSEASD KCQKEKRKTINGEDLLFAMGTLGFE EYVDPLKIYLHKYRE	19-109	82%
796	Os/G3398	REQDRFLPIANVSRIMKRALPANAK ISKDAKETVQECVSEFISFITGEASD KCQREKRKTINGDDLLWAMTTLGF EDYIDPLKLYLHKFRE	21-111	81%
	At/G1821 L1L NP_199578	REQDRFMPIANVIRIMRRILPAHAKI SDDSKETIQECVSEYISFITGEANER CQREQRKTITAEDVLWAMSKLGFD DYIEPLTLYLHRYRE	28-118	62%
	At/ AAC39488 LEC1	REQDQYMPIANVIRIMRKTLPSHAK ISDDAKETIQECVSEYISFVTGEANE RCQREQRKTITAEDILWAMSKLGFD NYVDPLTVFINRYRE	28-118	62%

Abbreviations: At *Arabidopsis thaliana*
Gm Glycine max
Os Oryza sativa
Zm Zea mays

G682 and the MYB-related transcription factors

G682 is a member of the MYB-related group of transcription factors. G682 and its related sequences were included a program to test the ability of these sequences to confer the same drought-related abiotic stress previously observed by us in 35S::G82 lines.

We first identified G682 as a putative transcription factor from the *Arabidopsis* BAC, AF007269, based on sequence similarity to other members of the MYB-related family within the conserved domain. To date, no functional data are available for this gene in the literature. The gene corresponds to

At4G01060, annotated by the *Arabidopsis* Genome initiative. G682 is one of a 5-member clade of related proteins that range in size from 75 to 112 amino acids. These proteins contain a single MYB repeat, which is not uncommon for plant MYB transcription factors. Information on gene function has been published for two of the genes in this clade, *CAPRICE* (*CPC*/G225) and *TRIPTYCHON* (*TRY*/G1816). Published information on gene function is not available for two additional members of the clade; G226 and G2718. our initial genomics program, members of the G682 clade were found to promote epidermal cell type alterations when overexpressed in *Arabidopsis*. These changes include both increased numbers of root hairs compared to wild type plants, as well as a reduction in trichome number. In addition, overexpression lines for all members of the clade showed a reduction in anthocyanin accumulation in response to stress, and enhanced tolerance to osmotic stress. In the case of *35S::G682* transgenic lines, an enhanced tolerance to high heat conditions was also observed. Given the phenotypic responses for G682 and its clade members, a sizeable number of clade members were included in the present drought-stress study. Table 2 summarizes the functional genomics data found in our investigation of G682 and its clade members.

Table 2: G682-clade experimental observations

	GID			
Observation	G226	G682	TRY (G1816)	G2718
Reduction in Trichome #	X	X	X	X
Increased Root Hair #	X	X	X	X
N Tolerance	X		X	X
Heat Tolerance		X		
Salt Tolerance	X			
Sugar response			X	

There are approximately 50 members of this family in *Arabidopsis*. The MYB-related DNA-binding domain contains approximately 50 amino acids with a series of highly conserved residues arranged with a characteristic spacing. The single-repeat MYB proteins do not contain a typical transcriptional activation domain and this suggests that they may function by interfering with the formation or activity of transcription factors or transcription factor complexes (Wada et al. (1997) *Science* 277: 1113-1116; Schellmann et al. (2002) *EMBO J.* 21: 5036-5046). In addition to the G682 clade, two well characterized transcription factors, *CIRCADIAN CLOCK ASSOCIATED1* (*CCA1*/G214) and *LATE ELONGATED HYPOCOTYL* (*LHY*/G680) represent two additional well-characterized MYB-related proteins that contain single MYB repeats (Wang et al. (1997) *Plant Cell* 9: 491-507; Schaffer et al. (1998) *Cell* 93: 1219-1229).

The difference in the phenotypic responses of the G682-clade overexpression lines (Table 2), along with the differences in the *CPC* (G225) and *TRY* (G1816) mutant phenotypes (Schellmann et al. (2002) *supra*), suggest that each of the five *Arabidopsis* genes in the clade have distinct but overlapping functions in the plant. In the case of *35S::G682* transgenic lines, an enhanced tolerance to high heat

conditions was observed. Heat can cause osmotic stress, and it is therefore consistent that these transgenic lines were also more tolerant to drought stress in a soil-based assay. Another common feature for four of five members of this clade is that they enhance performance under nitrogen-limiting conditions. 35S::G682 plants lacked this feature in the first round of screening, but given the high throughput nature of the genomics program, it is possible this phenotype would have been observed if a greater number of lines had been examined.

All of the genes in the *Arabidopsis* G682 clade reduced trichomes and increased root hairs when constitutively overexpressed (Table 2). It is unknown, however, whether the drought-tolerance phenotypes in these lines is related to the increase in root hairs on the root epidermis. Increasing root hair density may increase in absorptive surface area and increase in nitrate transporters that are normally found there. Alternatively, the *wer*, *ttg1* and *gl2* mutations, all of which increase root hair frequency, and have also been shown to cause ectopic stomate formation on the epidermis of hypocotyls. Thus, it is possible that *CPC* and *TRY* could be involved in the development, or regulation, of stomates as well (Hung et al. (1998) *Plant Physiol.* 117: 73-84, 1998; Berger et al. (1998) *Curr. Biol.* 8: 421-430; Lee and Schiefelbein (1999) *Cell* 99: 473-483). The *CPC* (G225) and *TRY* (G1816) proteins have not been reported to alter hypocotyl epidermal cell fate; however, ectopic expression may have resulted in an alteration in stomate production in this tissue. Since alterations in stomate production could alter plant water status, it should be examined in transgenics expressing the genes of the G682 clade, particularly in lines that show increased drought tolerance.

Interestingly, our data also suggest that G1816 (*TRY*) overexpression lines had a glucose sugar sensing phenotype. Several sugar sensing mutants have turned out to be allelic to ABA and ethylene mutants. This potentially implicates G1816 in hormone signaling.

As noted below, overexpression of a number of *Arabidopsis* and non-*Arabidopsis* the members of the G682 subclade of MYB-related transcription factors conferred increased abiotic stress tolerance in transgenic plants, as compared to non-transgenic plants of the same species a (i.e., non-transformed plant that did not overexpress these polypeptides).

Table 3 shows the polypeptides identified by SEQ ID NO; Gene ID (GID) No.; the transcription factor family to which the polypeptide belongs, and conserved MYB-related domains of the polypeptide. The first column shows the polypeptide SEQ ID NO; the second column the species and GID; the third column shows the MYB-related domain; the fourth column shows the amino acid coordinates of the conserved domain that were used to determine the percentage identity of that conserved domain to the MYB-related domain of G682,; and the fifth column shows the percentage identity to G682. The sequences are arranged in descending order of percentage identity to G682.

Table 3. Gene families and MYB-related domains

Polypeptide SEQ ID NO:	Species/ GID No., Accession No., or Identifier	MYB-related Domain	Amino Acid Coordinates for Percent Identity Determination	% ID MYB-related conserved domain of G682
148	At/G682	EWEVVNMSQEEEDLVSRMHKL VGDRWELIAGRIPGRT	27-63	100%
559	Os/G3393	AQNFVHFTEEEEDLVFRMHRLV GNRWELIAGRIPGRT	31-63	78%
2192	At/G2718	EWEEIAMAQEEEDLICRMYKLV GERWDLIAGRIPGRT	28-64	75%
1082	Os/G3392	AQNFVHFTEEEEDIVFRMHRLV GNRWELIAGRIPGRT	32-64	75%
1089	Zm/G3431	AQHLVDFTEAEEDLVSRMHRLV GNRWEIIAGRIPGRT	30-63	73%
1084	Gm/G3450	EWKVIHMSEQEEDLIRRMYKLV GDKWNLIAGRIPGRK	16-51	72%
2142	At/G1816	EWEFINMTEQEEDLIFRMYRLV GDRWDLIAGRVPGRQ	26-61	69%
1088	Gm/G3449	GSSKVEFSEDEETLIIRMYKLVG ERWSLIAGRIPGRT	26-58	69%
1087	Gm/G3448	GSSKVEFSEDEETLIIRMYKLVG ERWSIIAGRIPGRT	26-58	66%
38	At/G226	EWEFISMTEQEEDLISRMYRLV GNRWDLIAGRVVGRK	34-69	63%
1086	Gm/G3446	QVSDVEFSEAEEILIAMVYNLVG ERWSLIAGRIPGRT	26-58	60%
1083	Gm/G3445	QVSDVEFSEAEEILIAMVYNLVG ERWSLIAGRIPGRT	25-57	60%

Abbreviations: At *Arabidopsis thaliana*
Gm Glycine max
Os Oryza sativa
Zm Zea mays

G682 and its paralogs and orthologs are composed (almost entirely) of a single MYB-repeat DNA binding domain that is highly conserved across plant species. An alignment of the G682-like proteins from *Arabidopsis*, soybean, rice and corn that are being analyzed in this trait module is shown in Figures 3A and 3B. No function has been assigned to any of these MYB-related genes outside of *Arabidopsis*.

Because the G682 clade members are short proteins that are comprised almost exclusively of a DNA binding motif, it is likely that they function as repressors. This is consistent with in expression analyses indicating that *CPC* represses its own transcription as well as that of *WER* and *GL2* (Wada et al. (2002) *supra*; Lee and Schiefelbein (2002) *supra*). Repression may occur at the level of DNA binding through competition with other factors at target promoters, although repression via protein-protein interactions cannot be excluded.

The AP2 family, including the G47/G2133 and G1792 clades.

AP2 (APETALA2) and EREBPs (Ethylene-Responsive Element Binding Proteins) are the prototypic members of a family of transcription factors unique to plants, whose distinguishing characteristic is that they contain the so-called AP2 DNA-binding domain (for a review, see Riechmann and Meyerowitz (1998) *Biol. Chem.* 379: 633-646) . The AP2 domain was first recognized as a repeated motif within the *Arabidopsis thaliana* AP2 protein (Jofuku et al. (1994) *Plant Cell* 6: 1211-1225). Shortly afterwards, four DNA-binding proteins from tobacco were identified that interact with a sequence that is essential for the responsiveness of some promoters to the plant hormone ethylene, and were designated as *ethylene-responsive element binding proteins* (EREBPs; Ohme-Takagi et al. (1995) *Plant Cell* 7: 173-182). The DNA-binding domain of EREBP-2 was mapped to a region that was common to all four proteins (Ohme-Takagi et al (1995) *supra*), and that was found to be closely related to the AP2 domain (Weigel (1995) *Plant Cell* 7: 388-389) but that did not bear sequence similarity to previously known DNA-binding motifs.

AP2/EREBP genes form a large family, with many members known in several plant species (Okamuro et al. (1997) *Proc. Natl. Acad. Sci.* USA 94: 7076-7081; Riechmann and Meyerowitz (1998) *supra*). The number of AP2/EREBP genes in the *Arabidopsis thaliana* genome is approximately 145 (Riechmann et al. (2000) *Science* 290: 2105-2110). The APETALA2 class is characterized by the presence of two AP2 DNA binding domains, and contains 14 genes. The AP2/ERF is the largest subfamily, and includes 125 genes which are involved in abiotic (DREB subgroup) and biotic (ERF subgroup) stress responses and the RAV subgroup includes 6 genes which all have a B3 DNA binding domain in addition to the AP2 DNA binding domain (Kagaya et al. (1999) *Nucleic Acids Res.* 27: 470-478).

Arabidopsis AP2 is involved in the specification of sepal and petal identity through its activity as a homeotic gene that forms part of the combinatorial genetic mechanism of floral organ identity determination and it is also required for normal ovule and seed development (Bowman et al. (1991) *Development* 112: 1-20; Jofuku et al. (1994) *supra*). *Arabidopsis* ANT is required for ovule development and it also plays a role in floral organ growth (Elliott et al. (1996) *Plant Cell* 8: 155-168; Klucher et al. (1996) *Plant Cell* 8: 137-153). Finally, maize Gl15 regulates leaf epidermal cell identity (Moose et al. (1996) *Genes Dev.* 10: 3018-3027).

The attack of a plant by a pathogen may induce defense responses that lead to resistance to the invasion, and these responses are associated with transcriptional activation of defense-related genes, among them those encoding pathogenesis-related (PR) proteins. The involvement of EREBP-like genes in controlling the plant defense response is based on the observation that many PR gene promoters contain a short cis-acting element that mediates their responsiveness to ethylene (ethylene appears to be one of several signal molecules controlling the activation of defense responses). Tobacco EREBP-1, -2, -3, and -4, and tomato Pti4, Pti5 and Pti6 proteins have been shown to recognize such cis-acting elements

(Ohme-Takagi (1995) *supra*; Zhou et al. (1997) *EMBO J.* 16: 3207-3218). In addition, Pti4, Pti5, and Pti6 proteins have been shown to directly interact with Pto, a protein kinase that confers resistance against *Pseudomonas syringae* pv tomato (Zhou et al. (1997) *supra*). Plants are also challenged by adverse environmental conditions like cold or drought, and EREBP-like proteins appear to be involved in the responses to these abiotic stresses as well. COR (for cold-regulated) gene expression is induced during cold acclimation, the process by which plants increase their resistance to freezing in response to low unfreezing temperatures. The *Arabidopsis* EREBP-like gene CBF1 (Stockinger et al. (1997) *Proc. Natl. Acad. Sci.* USA 94: 1035-1040) is a regulator of the cold acclimation response, because ectopic expression of CBF1 in *Arabidopsis* transgenic plants induced COR gene expression in the absence of a cold stimulus, and the plant freezing tolerance was increased (Jaglo-Ottosen et al. (1998) *Science* 280: 104-106). Finally, another *Arabidopsis* EREBP-like gene, ABI4, is involved in abscisic acid (ABA) signal transduction, because abi4 mutants are insensitive to ABA (ABA is a plant hormone that regulates many agronomically important aspects of plant development; Finkelstein et al. (1998) *Plant Cell* 10: 1043-1054).

The SCR family, including the G922 clade.

The *SCARECROW* gene, which regulates an asymmetric cell division essential for proper radial organization of root cell layers, was isolated from *Arabidopsis thaliana* by screening a genomic library with sequences flanking a T-DNA insertion causing a "scarecrow" mutation (Di Laurenzio et al. (1996) *Cell* 86, 423-433). The gene product was tentatively described as a transcription factor based on the presence of homopolymeric stretches of several amino acids, the presence of a basic domain similar to that of the basic-leucine zipper family of transcription factors, and the presence of leucine heptad repeats. The presence of several *Arabidopsis* ESTs with gene products homologous to the *SCARECROW* gene were noted. The ability of the *SCARECROW* gene to complement the scarecrow mutation was also demonstrated (Malamy et al. (1997) *Plant J.* 12, 957-963).

More recently, the *SCARECROW* homologue *RGA*, which encodes a negative regulator of the gibberellin signal transduction pathway, was isolated from *Arabidopsis* by genomic subtraction (Silverstone et al. (1998) *Plant Cell* 10, 155-169). The *RGA* gene was shown to be expressed in many different tissues and the RGA protein was shown to be localized to the nucleus. The same gene was isolated by Truong (Truong et al. (1997) *FEBS Lett.* 410: 213-218) by identifying cDNA clones which complement a yeast nitrogen metabolism mutant, suggesting that RGA may be involved in regulating diverse metabolic processes. Another *SCARECROW* homologue designated *GAI*, which also is involved in gibberellin signaling processes, has been isolated by Peng (Peng et al. (1997) *Genes Dev.* 11, 3194-3205). Interestingly, *GAI* is the gene that initiated the Green Revolution. Peng et al. (Peng et al. (1999) Nature 6741, 256-261) have recently shown that maize *GAI* orthologs, when mutated, result in plants that are shorter, have increased seed yield, and are more resistant to damage by rain and wind than wild type

plants. Based on the inclusion of the GAI, RGA and SCR genes in this family, it has also been referred to as the GRAS family (Pysh et al. (1999) *Plant J* 18, 111-19).

The scarecrow gene family has 32 members in the *Arabidopsis* genome.

The NAC family, including the G2053 clade.

The NAC family is a group of transcription factors that share a highly conserved N-terminal domain of about 150 amino acids, designated the NAC domain (NAC stands for Petunia, NAM, and *Arabidopsis*, ATAF1, ATAF2 and CUC2). This is believed to be a novel domain that is present in both monocot and dicot plants but is absent from yeast and animal proteins. One hundred and twelve members of the NAC family have been identified in the *Arabidopsis* genome. The NAC class of proteins can be divided into at least two sub-families on the basis of amino acid sequence similarities within the NAC domain. One sub-family is built around the NAM and CUC2 (cup-shaped cotyledon) proteins whilst the other sub-family contains factors with a NAC domain similar to those of ATAF1 and ATAF2.

Thus far, little is known about the function of different NAC family members. This is surprising given that there are 113 members in *Arabidopsis*. However, NAM, CUC1 and CUC2 are thought to have vital roles in the regulation of embryo and flower development. In Petunia, *nam* mutant embryos fail to develop a shoot apical meristem (SAM) and have fused cotyledons. These mutants sometimes generate escape shoots that produce defective flowers with extra petals and fused organs. In *Arabidopsis*, the *cuc1* and *cuc2* mutations have somewhat similar effects, causing defects in SAM formation and the separation of cotyledons, sepals and stamens.

Although *nam* and *cuc* mutants exhibit comparable defects during embryogenesis, the penetrance of these phenotypes is much lower in *cuc* mutants. Functional redundancy of the CUC genes in *Arabidopsis* may explain this observation. In terms of the flower phenotype there are notable differences between *nam* and *cuc* mutants. Flowers of *cuc* mutants do not contain additional organs and the formation of sepals and stamens is most strongly affected. In *nam* mutants, by contrast, the flowers do carry additional organs and petal formation is more markedly affected than that of other floral organs. These apparent differences might be explained in two ways: the NAM and CUC proteins have been recruited into different roles in development of *Arabidopsis* and Petunia flowers. Alternatively, the proteins could share a common function between the two species, with the different mutant floral phenotypes arising from variations in the way other genes (that participate in the same developmental processes) are affected by defects in NAM or CUC.

A further gene from this family, NAP (NAC-like activated by AP3/PI) is also involved in flower development and is thought to influence the transition between cell division and cell expansion in stamens and petals. Overall, then, the NAC proteins mainly appear to regulate developmental processes.

Producing Polypeptides

The polynucleotides of the invention include sequences that encode transcription factors and transcription factor homolog polypeptides and sequences complementary thereto, as well as unique fragments of coding sequence, or sequence complementary thereto. Such polynucleotides can be, e.g., DNA or RNA, e.g., mRNA, cRNA, synthetic RNA, genomic DNA, cDNA synthetic DNA, oligonucleotides, etc. The polynucleotides are either double-stranded or single-stranded, and include either, or both sense (i.e., coding) sequences and antisense (i.e., non-coding, complementary) sequences. The polynucleotides include the coding sequence of a transcription factor, or transcription factor homolog polypeptide, in isolation, in combination with additional coding sequences (e.g., a purification tag, a localization signal, as a fusion-protein, as a pre-protein, or the like), in combination with non-coding sequences (e.g., introns or inteins, regulatory elements such as promoters, enhancers, terminators, and the like), and/or in a vector or host environment in which the polynucleotide encoding a transcription factor or transcription factor homolog polypeptide is an endogenous or exogenous gene.

A variety of methods exist for producing the polynucleotides of the invention. Procedures for identifying and isolating DNA clones are well known to those of skill in the art, and are described in, e.g., Berger and Kimmel, Guide to Molecular Cloning Techniques, Methods in Enzymology, vol. 152 Academic Press, Inc., San Diego, CA ("Berger"); Sambrook et al. (1989) Molecular Cloning - A Laboratory Manual (2nd Ed.), Vol. 1-3, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., and Current Protocols in Molecular Biology, Ausubel et al. eds., Current Protocols, a joint venture between Greene Publishing Associates, Inc. and John Wiley & Sons, Inc., (supplemented through 2000) ("Ausubel").

Alternatively, polynucleotides of the invention, can be produced by a variety of in vitro amplification methods adapted to the present invention by appropriate selection of specific or degenerate primers. Examples of protocols sufficient to direct persons of skill through in vitro amplification methods, including the polymerase chain reaction (PCR) the ligase chain reaction (LCR), Qbeta-replicase amplification and other RNA polymerase mediated techniques (e.g., NASBA), e.g., for the production of the homologous nucleic acids of the invention are found in Berger (supra), Sambrook (supra), and Ausubel (supra), as well as Mullis et al. (1987) PCR Protocols A Guide to Methods and Applications (Innis et al. eds) Academic Press Inc. San Diego, CA (1990) (Innis). Improved methods for cloning in vitro amplified nucleic acids are described in Wallace et al. U.S. Pat. No. 5,426,039. Improved methods for amplifying large nucleic acids by PCR are summarized in Cheng et al. (1994) *Nature* 369: 684-685 and the references cited therein, in which PCR amplicons of up to 40kb are generated. One of skill will appreciate that essentially any RNA can be converted into a double stranded DNA suitable for restriction digestion, PCR expansion and sequencing using reverse transcriptase and a polymerase. See, e.g., Ausubel, Sambrook and Berger, all *supra*.

Alternatively, polynucleotides and oligonucleotides of the invention can be assembled from fragments produced by solid-phase synthesis methods. Typically, fragments of up to approximately 100

bases are individually synthesized and then enzymatically or chemically ligated to produce a desired sequence, e.g., a polynucleotide encoding all or part of a transcription factor. For example, chemical synthesis using the phosphoramidite method is described, e.g., by Beaucage et al. (1981) *Tetrahedron Letters* 22: 1859-1869; and Matthes et al. (1984) *EMBO J.* 3: 801-805. According to such methods, oligonucleotides are synthesized, purified, annealed to their complementary strand, ligated and then optionally cloned into suitable vectors. And if so desired, the polynucleotides and polypeptides of the invention can be custom ordered from any of a number of commercial suppliers.

Homologous Sequences

Sequences homologous, i.e., that share significant sequence identity or similarity, to those provided in the Sequence Listing, derived from *Arabidopsis thaliana* or from other plants of choice, are also an aspect of the invention. Homologous sequences can be derived from any plant including monocots and dicots and in particular agriculturally important plant species, including but not limited to, crops such as soybean, wheat, corn (maize), potato, cotton, rice, rape, oilseed rape (including canola), sunflower, alfalfa, clover, sugarcane, and turf; or fruits and vegetables, such as banana, blackberry, blueberry, strawberry, and raspberry, cantaloupe, carrot, cauliflower, coffee, cucumber, eggplant, grapes, honeydew, lettuce, mango, melon, onion, papaya, peas, peppers, pineapple, pumpkin, spinach, squash, sweet corn, tobacco, tomato, tomatillo, watermelon, rosaceous fruits (such as apple, peach, pear, cherry and plum) and vegetable brassicas (such as broccoli, cabbage, cauliflower, Brussels sprouts, and kohlrabi). Other crops, including fruits and vegetables, whose phenotype can be changed and which comprise homologous sequences include barley; rye; millet; sorghum; currant; avocado; citrus fruits such as oranges, lemons, grapefruit and tangerines, artichoke, cherries; nuts such as the walnut and peanut; endive; leek; roots such as arrowroot, beet, cassava, turnip, radish, yam, and sweet potato; and beans. The homologous sequences may also be derived from woody species, such pine, poplar and eucalyptus, or mint or other labiates. In addition, homologous sequences may be derived from plants that are evolutionarily-related to crop plants, but which may not have yet been used as crop plants. Examples include deadly nightshade (*Atropa belladona*), related to tomato; jimson weed (Datura strommium), related to peyote; and teosinte (*Zea* species), related to corn (maize).

Orthologs and Paralogs

Homologous sequences as described above can comprise orthologous or paralogous sequences. Several different methods are known by those of skill in the art for identifying and defining these functionally homologous sequences. Three general methods for defining orthologs and paralogs are described; an ortholog or paralog, including equivalogs, may be identified by one or more of the methods described below.

Orthologs and paralogs are evolutionarily related genes that have similar sequence and similar functions. Orthologs are structurally related genes in different species that are derived by a speciation

event. Paralogs are structurally related genes within a single species that are derived by a duplication event.

Within a single plant species, gene duplication may cause two copies of a particular gene, giving rise to two or more genes with similar sequence and often similar function known as paralogs. A paralog is therefore a similar gene formed by duplication within the same species. Paralogs typically cluster together or in the same clade (a group of similar genes) when a gene family phylogeny is analyzed using programs such as CLUSTAL (Thompson et al. (1994) Nucleic Acids Res. 22: 4673-4680; Higgins et al. (1996) Methods Enzymol. 266: 383-402). Groups of similar genes can also be identified with pair-wise BLAST analysis (Feng and Doolittle (1987) *J. Mol. Evol.* 25: 351-360). For example, a clade of very similar MADS domain transcription factors from *Arabidopsis* all share a common function in flowering time (Ratcliffe et al. (2001) *Plant Physiol.* 126: 122-132), and a group of very similar AP2 domain transcription factors from *Arabidopsis* are involved in tolerance of plants to freezing (Gilmour et al. (1998) Plant J. 16: 433-442). Analysis of groups of similar genes with similar function that fall within one clade can yield sub-sequences that are particular to the clade. These sub-sequences, known as consensus sequences, can not only be used to define the sequences within each clade, but define the functions of these genes; genes within a clade may contain paralogous sequences, or orthologous sequences that share the same function (see also, for example, Mount (2001), in Bioinformatics: Sequence and Genome Analysis, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, page 543.)

Speciation, the production of new species from a parental species, can also give rise to two or more genes with similar sequence and similar function. These genes, termed orthologs, often have an identical function within their host plants and are often interchangeable between species without losing function. Because plants have common ancestors, many genes in any plant species will have a corresponding orthologous gene in another plant species. Once a phylogenic tree for a gene family of one species has been constructed using a program such as CLUSTAL (Thompson et al. (1994) Nucleic Acids Res. 22: 4673-4680; Higgins et al. (1996) *supra*) potential orthologous sequences can be placed into the phylogenetic tree and their relationship to genes from the species of interest can be determined. Orthologous sequences can also be identified by a reciprocal BLAST strategy. Once an orthologous sequence has been identified, the function of the ortholog can be deduced from the identified function of the reference sequence.

Transcription factor gene sequences are conserved across diverse eukaryotic species lines (Goodrich et al. (1993) *Cell* 75: 519-530; Lin et al. (1991) *Nature* 353: 569-571; Sadowski et al. (1988) *Nature* 335: 563-564).et al. Plants are no exception to this observation; diverse plant species possess transcription factors that have similar sequences and functions.

Orthologous genes from different organisms have highly conserved functions, and very often essentially identical functions (Lee et al. (2002) *Genome Res.* 12: 493-502; Remm et al. (2001) *J. Mol. Biol.* 314: 1041-1052). Paralogous genes, which have diverged through gene duplication, may retain

similar functions of the encoded proteins. In such cases, paralogs can be used interchangeably with respect to certain embodiments of the instant invention (for example, transgenic expression of a coding sequence). An example of such highly related paralogs is the CBF family, with three well-defined members in *Arabidopsis* and at least one ortholog in *Brassica napus* (SEQ ID NOs: 1956, 1958, 1960, or 2204, respectively), all of which control pathways involved in both freezing and drought stress (Gilmour et al. (1998) *Plant J.* 16: 433-442; Jaglo et al. (1998) *Plant Physiol.* 127: 910-917).

The following references represent a small sampling of the many studies that demonstrate that conserved transcription factor genes from diverse species are likely to function similarly (i.e., regulate similar target sequences and control the same traits), and that transcription factors may be transformed into diverse species to confer or improve traits.

(1) The *Arabidopsis* NPR1 gene regulates systemic acquired resistance (SAR); over-expression of NPR1 leads to enhanced resistance in *Arabidopsis*. When either *Arabidopsis* NPR1 or the rice NPR1 ortholog was overexpressed in rice (which, as a monocot, is diverse from *Arabidopsis*), challenge with the rice bacterial blight pathogen *Xanthomonas oryzae* pv. Oryzae, the transgenic plants displayed enhanced resistance (Chern et al. (2001) *Plant J.* 27: 101-113). NPR1 acts through activation of expression of transcription factor genes, such as TGA2 (Fan and Dong (2002) *Plant Cell* 14: 1377-1389).

(2) E2F genes are involved in transcription of plant genes for proliferating cell nuclear antigen (PCNA). Plant E2Fs share a high degree of similarity in amino acid sequence between monocots and dicots, and are even similar to the conserved domains of the animal E2Fs. Such conservation indicates a functional similarity between plant and animal E2Fs. E2F transcription factors that regulate meristem development act through common cis-elements, and regulate related (PCNA) genes (Kosugi and Ohashi, (2002) *Plant J.* 29: 45-59).

(3) The ABI5 gene (abscisic acid (ABA) insensitive 5) encodes a basic leucine zipper factor required for ABA response in the seed and vegetative tissues. Co-transformation experiments with ABI5 cDNA constructs in rice protoplasts resulted in specific transactivation of the ABA-inducible wheat, *Arabidopsis*, bean, and barley promoters. These results demonstrate that sequentially similar ABI5 transcription factors are key targets of a conserved ABA signaling pathway in diverse plants. (Gampala et al. (2001) *J. Biol. Chem.* 277: 1689-1694).

(4) Sequences of three *Arabidopsis* GAMYB-like genes were obtained on the basis of sequence similarity to GAMYB genes from barley, rice, and *L. temulentum*. These three *Arabadopsis* genes were determined to encode transcription factors (AtMYB33, AtMYB65, and AtMYB101) and could substitute for a barley GAMYB and control alpha-amylase expression (Gocal et al. (2001) *Plant Physiol.* 127: 1682-1693).

(5) The floral control gene LEAFY from *Arabidopsis* can dramatically accelerate flowering in numerous dictoyledonous plants. Constitutive expression of *Arabidopsis* LEAFY also caused early flowering in transgenic rice (a monocot), with a heading date that was 26-34 days earlier than that of

wild-type plants. These observations indicate that floral regulatory genes from *Arabidopsis* are useful tools for heading date improvement in cereal crops (He et al. (2000) *Transgenic Res.* 9: 223-227).

(6) Bioactive gibberellins (GAs) are essential endogenous regulators of plant growth. GA signaling tends to be conserved across the plant kingdom. GA signaling is mediated via GAI, a nuclear member of the GRAS family of plant transcription factors. *Arabidopsis* GAI has been shown to function in rice to inhibit gibberellin response pathways (Fu et al. (2001) *Plant Cell* 13: 1791-1802).

(7) The *Arabidopsis* gene SUPERMAN (SUP), encodes a putative transcription factor that maintains the boundary between stamens and carpels. By over-expressing *Arabidopsis* SUP in rice, the effect of the gene's presence on whorl boundaries was shown to be conserved. This demonstrated that SUP is a conserved regulator of floral whorl boundaries and affects cell proliferation (Nandi et al. (2000) *Curr. Biol.* 10: 215-218).

(8) Maize, petunia and *Arabidopsis* myb transcription factors that regulate flavonoid biosynthesis are very genetically similar and affect the same trait in their native species, therefore sequence and function of these myb transcription factors correlate with each other in these diverse species (Borevitz et al. (2000) *Plant Cell* 12: 2383-2394).

(9) Wheat reduced height-1 (Rht-B1/Rht-D1) and maize dwarf-8 (d8) genes are orthologs of the *Arabidopsis* gibberellin insensitive (GAI) gene. Both of these genes have been used to produce dwarf grain varieties that have improved grain yield. These genes encode proteins that resemble nuclear transcription factors and contain an SH2-like domain, indicating that phosphotyrosine may participate in gibberellin signaling. Transgenic rice plants containing a mutant GAI allele from *Arabidopsis* have been shown to produce reduced responses to gibberellin and are dwarfed, indicating that mutant GAI orthologs could be used to increase yield in a wide range of crop species (Peng et al. (1999) *Nature* 400: 256-261).

Transcription factors that are homologous to the listed sequences will typically share, in at least one conserved domain, at least about 70% amino acid sequence identity, and with regard to zinc finger transcription factors, at least about 50% amino acid sequence identity. More closely related transcription factors can share at least about 70%,or about 75% or about 80% or about 90% or about 95% or about 98% or more sequence identity with the listed sequences, or with the listed sequences but excluding or outside a known consensus sequence or consensus DNA-binding site, or with the listed sequences excluding one or all conserved domain. Factors that are most closely related to the listed sequences share, e.g., at least about 85%, about 90% or about 95% or more % sequence identity to the listed sequences, or to the listed sequences but excluding or outside a known consensus sequence or consensus DNA-binding site or outside one or all conserved domain. At the nucleotide level, the sequences will typically share at least about 40% nucleotide sequence identity, preferably at least about 50%, about 60%, about 70% or about 80% sequence identity, and more preferably about 85%, about 90%, about 95% or about 97% or more sequence identity to one or more of the listed sequences, or to a listed sequence but excluding or outside a known consensus sequence or consensus DNA-binding site, or outside one or all conserved domain. The degeneracy of the genetic code enables major variations in the nucleotide sequence of a

polynucleotide while maintaining the amino acid sequence of the encoded protein. Conserved domains within a transcription factor family may exhibit a higher degree of sequence homology, such as at least 65% amino acid sequence identity including conservative substitutions, and preferably at least 80% sequence identity, and more preferably at least 85%, or at least about 86%, or at least about 87%, or at least about 88%, or at least about 90%, or at least about 95%, or at least about 98% sequence identity. Transcription factors that are homologous to the listed sequences should share at least 30%, or at least about 60%, or at least about 75%, or at least about 80%, or at least about 90%, or at least about 95% amino acid sequence identity over the entire length of the polypeptide or the homolog.

Percent identity can be determined electronically, e.g., by using the MEGALIGN program (DNASTAR, Inc. Madison, Wis.). The MEGALIGN program can create alignments between two or more sequences according to different methods, for example, the clustal method. (See, for example, Higgins and Sharp (1988) *Gene* 73: 237-244.) The clustal algorithm groups sequences into clusters by examining the distances between all pairs. The clusters are aligned pairwise and then in groups. Other alignment algorithms or programs may be used, including FASTA, BLAST, or ENTREZ, FASTA and BLAST, and which may be used to calculate percent similarity. These are available as a part of the GCG sequence analysis package (University of Wisconsin, Madison, Wis.), and can be used with or without default settings. ENTREZ is available through the National Center for Biotechnology Information. In one embodiment, the percent identity of two sequences can be determined by the GCG program with a gap weight of 1, e.g., each amino acid gap is weighted as if it were a single amino acid or nucleotide mismatch between the two sequences (see USPN 6,262,333).

Other techniques for alignment are described in Doolittle, R. F. (1996) Methods in Enzymology: Computer Methods for Macromolecular Sequence Analysis, vol. 266, Academic Press, Orlando, FL, USA. Preferably, an alignment program that permits gaps in the sequence is utilized to align the sequences. The Smith-Waterman is one type of algorithm that permits gaps in sequence alignments (see Shpaer (1997) *Methods Mol. Biol.* 70: 173-187). Also, the GAP program using the Needleman and Wunsch alignment method can be utilized to align sequences. An alternative search strategy uses MPSRCH software, which runs on a MASPAR computer. MPSRCH uses a Smith-Waterman algorithm to score sequences on a massively parallel computer. This approach improves ability to pick up distantly related matches, and is especially tolerant of small gaps and nucleotide sequence errors. Nucleic acid-encoded amino acid sequences can be used to search both protein and DNA databases.

The percentage similarity between two polypeptide sequences, e.g., sequence A and sequence B, is calculated by dividing the length of sequence A, minus the number of gap residues in sequence A, minus the number of gap residues in sequence B, into the sum of the residue matches between sequence A and sequence B, times one hundred. Gaps of low or of no similarity between the two amino acid sequences are not included in determining percentage similarity. Percent identity between polynucleotide sequences can also be counted or calculated by other methods known in the art, e.g., the Jotun Hein method. (See, e.g., Hein (1990) *Methods Enzymol.* 183: 626-645.) Identity between sequences can also

be determined by other methods known in the art, e.g., by varying hybridization conditions (see US Patent Application No. 20010010913).

The percent identity between two conserved domains of a transcription factor DNA-binding domain consensus polypeptide sequence can be as low as 16%, as exemplified in the case of GATA1 family of eukaryotic Cys_2/Cys_2-type zinc finger transcription factors. The DNA-binding domain consensus polypeptide sequence of the GATA1 family is $CX_2CX_{17}CX_2C$, where X is any amino acid residue. (See, for example, Takatsuji, *supra*.) Other examples of such conserved consensus polypeptide sequences with low overall percent sequence identity are well known to those of skill in the art.

Thus, the invention provides methods for identifying a sequence similar or paralogous or orthologous or homologous to one or more polynucleotides as noted herein, or one or more target polypeptides encoded by the polynucleotides, or otherwise noted herein and may include linking or associating a given plant phenotype or gene function with a sequence. In the methods, a sequence database is provided (locally or across an internet or intranet) and a query is made against the sequence database using the relevant sequences herein and associated plant phenotypes or gene functions.

In addition, one or more polynucleotide sequences or one or more polypeptides encoded by the polynucleotide sequences may be used to search against a BLOCKS (Bairoch et al. (1997) *Nucleic Acids Res.* 25: 217-221), PFAM, and other databases which contain previously identified and annotated motifs, sequences and gene functions. Methods that search for primary sequence patterns with secondary structure gap penalties (Smith et al. (1992) *Protein Engineering* 5: 35-51) as well as algorithms such as Basic Local Alignment Search Tool (BLAST; Altschul (1993) *J. Mol. Evol.* 36: 290-300; Altschul et al. (1990) *supra*), BLOCKS (Henikoff and Henikoff (1991) *Nucleic Acids Res.* 19: 6565-6572), Hidden Markov Models (HMM; Eddy (1996) *Curr. Opin. Str. Biol.* 6: 361-365; Sonnhammer et al. (1997) *Proteins* 28: 405-420), and the like, can be used to manipulate and analyze polynucleotide and polypeptide sequences encoded by polynucleotides. These databases, algorithms and other methods are well known in the art and are described in Ausubel et al. (1997; Short Protocols in Molecular Biology, John Wiley & Sons, New York, NY, unit 7.7) and in Meyers (1995; Molecular Biology and Biotechnology, Wiley VCH, New York, NY, p 856-853).

Furthermore, methods using manual alignment of sequences similar or homologous to one or more polynucleotide sequences or one or more polypeptides encoded by the polynucleotide sequences may be used to identify regions of similarity and conserved domains. Such manual methods are well-known of those of skill in the art and can include, for example, comparisons of tertiary structure between a polypeptide sequence encoded by a polynucleotide which comprises a known function with a polypeptide sequence encoded by a polynucleotide sequence which has a function not yet determined. Such examples of tertiary structure may comprise predicted alpha helices, beta-sheets, amphipathic helices, leucine zipper motifs, zinc finger motifs, proline-rich regions, cysteine repeat motifs, and the like.

Orthologs and paralogs of presently disclosed transcription factors may be cloned using compositions provided by the present invention according to methods well known in the art. cDNAs can be cloned using mRNA from a plant cell or tissue that expresses one of the present transcription factors. Appropriate mRNA sources may be identified by interrogating Northern blots with probes designed from the present transcription factor sequences, after which a library is prepared from the mRNA obtained from a positive cell or tissue. Transcription factor-encoding cDNA is then isolated using, for example, PCR, using primers designed from a presently disclosed transcription factor gene sequence, or by probing with a partial or complete cDNA or with one or more sets of degenerate probes based on the disclosed sequences. The cDNA library may be used to transform plant cells. Expression of the cDNAs of interest is detected using, for example, methods disclosed herein such as microarrays, Northern blots, quantitative PCR, or any other technique for monitoring changes in expression. Genomic clones may be isolated using similar techniques to those.

In addition to the Sequences listed in the Sequence Listing, the invention encompasses isolated nucleotide sequences that are sequentially and structurally similar to G481, G482, G485, and G682, SEQ ID NO: 87, 89, 2009, and 147, respectively, and function in a plant in a manner similar to G481, G482, G485 and G867 by regulating abiotic stress tolerance.

The nucleotide sequences of the G482 subclade of the non-LEC1-like clade of proteins of the L1L-related CCAAT transcription factor family, including G481, share at least 81% identity in their B domains with the B domain of G481 (Table 1). Sequences outside of this subclade, including L1L (NP_199578) and LEC1 (AAC39488) share significantly less identity with G481 (Table 1), are phylogenetically distinct from the members of this subclade (Figures 6 and 7), and appear to function in embryonic development rather than in abiotic stress tolerance, as noted above.

Since the members of this subclade are phylogenetically related, sequentially similar and a representative number from diverse plant species have been shown to regulate abiotic stress tolerance, one skilled in the art would predict that other similar, phylogenetically related sequences would also regulate abiotic stress tolerance.

Similar to the G481 subclade, G682 and similar sequences in the G682 subclade of MYB-related transcription factors are phylogenetically related, sequentially similar and a representative number from diverse plant species have been shown to regulate abiotic stress tolerance. This would prompt one skilled in the art to draw similar conclusions regarding the regulation of these sequences of abiotic stress tolerance. A representative number of the members of this clade, including sequences derived from diverse non-*Arabidopsis* species, have been shown to confer abiotic stress tolerance when overexpressed. These sequences have been shown to share 60% identity in their MYB-related domains (Table 3).

Identifying Polynucleotides or Nucleic Acids by Hybridization

Polynucleotides homologous to the sequences illustrated in the Sequence Listing and tables can be identified, e.g., by hybridization to each other under stringent or under highly stringent conditions.

Single stranded polynucleotides hybridize when they associate based on a variety of well characterized physical-chemical forces, such as hydrogen bonding, solvent exclusion, base stacking and the like. The stringency of a hybridization reflects the degree of sequence identity of the nucleic acids involved, such that the higher the stringency, the more similar are the two polynucleotide strands. Stringency is influenced by a variety of factors, including temperature, salt concentration and composition, organic and non-organic additives, solvents, etc. present in both the hybridization and wash solutions and incubations (and number thereof), as described in more detail in the references cited above.

Encompassed by the invention are polynucleotide sequences that are capable of hybridizing to the claimed polynucleotide sequences, including any of the transcription factor polynucleotides within the Sequence Listing, and fragments thereof under various conditions of stringency (See, for example, Wahl and Berger (1987) *Methods Enzymol.* 152: 399-407; and Kimmel (1987) *Methods Enzymol.* 152: 507-511). In addition to the nucleotide sequences listed in Tables 7 and 8, full length cDNA, orthologs, and paralogs of the present nucleotide sequences may be identified and isolated using well-known methods. The cDNA libraries orthologs, and paralogs of the present nucleotide sequences may be screened using hybridization methods to determine their utility as hybridization target or amplification probes.

With regard to hybridization, conditions that are highly stringent, and means for achieving them, are well known in the art. See, for example, Sambrook et al. (1989) "*Molecular Cloning: A Laboratory Manual*" (2nd ed., Cold Spring Harbor Laboratory); Berger and Kimmel, eds., (1987) "Guide to Molecular Cloning Techniques", In *Methods in Enzymology*:152: 467-469; and Anderson and Young (1985) "Quantitative Filter Hybridisation." In: Hames and Higgins, ed., Nucleic Acid Hybridisation, A Practical Approach. Oxford, IRL Press, 73–111.

Stability of DNA duplexes is affected by such factors as base composition, length, and degree of base pair mismatch. Hybridization conditions may be adjusted to allow DNAs of different sequence relatedness to hybridize. The melting temperature (T_m) is defined as the temperature when 50% of the duplex molecules have dissociated into their constituent single strands. The melting temperature of a perfectly matched duplex, where the hybridization buffer contains formamide as a denaturing agent, may be estimated by the following equation:

(1) DNA-DNA: $T_m(^\circ C)=81.5+16.6(\log [Na+])+0.41(\%\ G+C)-0.62(\%\ \text{formamide})-500/L$

(2) DNA-RNA: $T_m(^\circ C)=79.8+18.5(\log [Na+])+0.58(\%\ G+C)+0.12(\%G+C)^2-0.5(\%\ \text{formamide})-820/L$

(3) RNA-RNA: $T_m(^\circ C)=79.8+18.5(\log [Na+])+0.58(\%\ G+C)+0.12(\%G+C)^2-0.35(\%\ \text{formamide})-820/L$

where L is the length of the duplex formed, [Na+] is the molar concentration of the sodium ion in the hybridization or washing solution, and % G+C is the percentage of (guanine+cytosine) bases in the

hybrid. For imperfectly matched hybrids, approximately 1° C is required to reduce the melting temperature for each 1-% mismatch.

Hybridization experiments are generally conducted in a buffer of pH between 6.8 to 7.4, although the rate of hybridization is nearly independent of pH at ionic strengths likely to be used in the hybridization buffer (Anderson et al. (1985) *supra*). In addition, one or more of the following may be used to reduce non-specific hybridization: sonicated salmon sperm DNA or another non-complementary DNA, bovine serum albumin, sodium pyrophosphate, sodium dodecylsulfate (SDS), polyvinyl-pyrrolidone, ficoll and Denhardt's solution. Dextran sulfate and polyethylene glycol 6000 act to exclude DNA from solution, thus raising the effective probe DNA concentration and the hybridization signal within a given unit of time. In some instances, conditions of even greater stringency may be desirable or required to reduce non-specific and/or background hybridization. These conditions may be created with the use of higher temperature, lower ionic strength and higher concentration of a denaturing agent such as formamide.

Stringency conditions can be adjusted to screen for moderately similar fragments such as homologous sequences from distantly related organisms, or to highly similar fragments such as genes that duplicate functional enzymes from closely related organisms. The stringency can be adjusted either during the hybridization step or in the post-hybridization washes. Salt concentration, formamide concentration, hybridization temperature and probe lengths are variables that can be used to alter stringency (as described by the formula above). As a general guidelines high stringency is typically performed at T_m-5oC to T_m-20oC, moderate stringency at T_m-20oC to T_m-35oC and low stringency at T_m-35oC to T_m-50oC for duplex >150 base pairs. Hybridization may be performed at low to moderate stringency (25-50oC below T_m), followed by post-hybridization washes at increasing stringencies. Maximum rates of hybridization in solution are determined empirically to occur at T_m-25oC for DNA-DNA duplex and T_m-15oC for RNA-DNA duplex. Optionally, the degree of dissociation may be assessed after each wash step to determine the need for subsequent, higher stringency wash steps.

High stringency conditions may be used to select for nucleic acid sequences with high degrees of identity to the disclosed sequences. An example of stringent hybridization conditions obtained in a filter-based method such as a Southern or northern blot for hybridization of complementary nucleic acids that have more than 100 complementary residues is about 5°C to 20°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. Conditions used for hybridization may include about 0.02 M to about 0.15 M sodium chloride, about 0.5% to about 5% casein, about 0.02% SDS or about 0.1% N-laurylsarcosine, about 0.001 M to about 0.03 M sodium citrate, at hybridization temperatures between about 50° C and about 70° C. More preferably, high stringency conditions are about 0.02 M sodium chloride, about 0.5% casein, about 0.02% SDS, about 0.001 M sodium citrate, at a temperature of about 50° C. Nucleic acid molecules that hybridize under stringent conditions will typically hybridize to a probe based on either the entire DNA molecule or selected portions, e.g., to a unique subsequence, of the DNA.

Stringent salt concentration will ordinarily be less than about 750 mM NaCl and 75 mM trisodium citrate. Increasingly stringent conditions may be obtained with less than about 500 mM NaCl and 50 mM trisodium citrate, to even greater stringency with less than about 250 mM NaCl and 25 mM trisodium citrate. Low stringency hybridization can be obtained in the absence of organic solvent, e.g., formamide, whereas high stringency hybridization may be obtained in the presence of at least about 35% formamide, and more preferably at least about 50% formamide. Stringent temperature conditions will ordinarily include temperatures of at least about 30° C, more preferably of at least about 37° C, and most preferably of at least about 42° C with formamide present. Varying additional parameters, such as hybridization time, the concentration of detergent, e.g., sodium dodecyl sulfate (SDS) and ionic strength, are well known to those skilled in the art. Various levels of stringency are accomplished by combining these various conditions as needed. In a preferred embodiment, hybridization will occur at 30° C in 750 mM NaCl, 75 mM trisodium citrate, and 1% SDS. In a more preferred embodiment, hybridization will occur at 37° C in 500 mM NaCl, 50 mM trisodium citrate, 1% SDS, 35% formamide. In a most preferred embodiment, hybridization will occur at 42° C in 250 mM NaCl, 25 mM trisodium citrate, 1% SDS, 50% formamide. Useful variations on these conditions will be readily apparent to those skilled in the art.

The washing steps that follow hybridization may also vary in stringency; the post-hybridization wash steps primarily determine hybridization specificity, with the most critical factors being temperature and the ionic strength of the final wash solution. Wash stringency can be increased by decreasing salt concentration or by increasing temperature. Stringent salt concentration for the wash steps will preferably be less than about 30 mM NaCl and 3 mM trisodium citrate, and most preferably less than about 15 mM NaCl and 1.5 mM trisodium citrate. For example, the wash conditions may be under conditions of 0.1 x SSC to 2.0 x SSC and 0.1% SDS at 50-65° C, with, for example, two steps of 10 - 30 min. One example of stringent wash conditions includes about 2.0 x SSC, 0.1% SDS at 65° C and washing twice, each wash step being about 30 min. A higher stringency wash is about 0.2 x SSC, 0.1% SDS at 65° C and washing twice for 30 min. A still higher stringency wash is about 0.1 x SSC, 0.1% SDS at 65° C and washing twice for 30 min. The temperature for the wash solutions will ordinarily be at least about 25° C, and for greater stringency at least about 42° C. Hybridization stringency may be increased further by using the same conditions as in the hybridization steps, with the wash temperature raised about 3° C to about 5° C, and stringency may be increased even further by using the same conditions except the wash temperature is raised about 6° C to about 9° C. For identification of less closely related homolog, wash steps may be performed at a lower temperature, e.g., 50° C.

An example of a low stringency wash step employs a solution and conditions of at least 25° C in 30 mM NaCl, 3 mM trisodium citrate, and 0.1% SDS over 30 min. Greater stringency may be obtained at 42° C in 15 mM NaCl, with 1.5 mM trisodium citrate, and 0.1% SDS over 30 min. Even higher stringency wash conditions are obtained at 65° C -68° C in a solution of 15 mM NaCl, 1.5 mM trisodium citrate, and 0.1% SDS. Wash procedures will generally employ at least two final wash steps. Additional

variations on these conditions will be readily apparent to those skilled in the art (see, for example, U.S. Patent Application No. 20010010913).

Stringency conditions can be selected such that an oligonucleotide that is perfectly complementary to the coding oligonucleotide hybridizes to the coding oligonucleotide with at least about a 5-10x higher signal to noise ratio than the ratio for hybridization of the perfectly complementary oligonucleotide to a nucleic acid encoding a transcription factor known as of the filing date of the application. It may be desirable to select conditions for a particular assay such that a higher signal to noise ratio, that is, about 15x or more, is obtained. Accordingly, a subject nucleic acid will hybridize to a unique coding oligonucleotide with at least a 2x or greater signal to noise ratio as compared to hybridization of the coding oligonucleotide to a nucleic acid encoding known polypeptide. The particular signal will depend on the label used in the relevant assay, e.g., a fluorescent label, a colorimetric label, a radioactive label, or the like. Labeled hybridization or PCR probes for detecting related polynucleotide sequences may be produced by oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide.

Identifying Polynucleotides or Nucleic Acids with Expression Libraries

In addition to hybridization methods, transcription factor homolog polypeptides can be obtained by screening an expression library using antibodies specific for one or more transcription factors. With the provision herein of the disclosed transcription factor, and transcription factor homolog nucleic acid sequences, the encoded polypeptide(s) can be expressed and purified in a heterologous expression system (e.g., *E. coli*) and used to raise antibodies (monoclonal or polyclonal) specific for the polypeptide(s) in question. Antibodies can also be raised against synthetic peptides derived from transcription factor, or transcription factor homolog, amino acid sequences. Methods of raising antibodies are well known in the art and are described in Harlow and Lane (1988), Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory, New York. Such antibodies can then be used to screen an expression library produced from the plant from which it is desired to clone additional transcription factor homologs, using the methods described above. The selected cDNAs can be confirmed by sequencing and enzymatic activity.

Sequence Variations

It will readily be appreciated by those of skill in the art, that any of a variety of polynucleotide sequences are capable of encoding the transcription factors and transcription factor homolog polypeptides of the invention. Due to the degeneracy of the genetic code, many different polynucleotides can encode identical and/or substantially similar polypeptides in addition to those sequences illustrated in the Sequence Listing. Nucleic acids having a sequence that differs from the sequences shown in the Sequence Listing, or complementary sequences, that encode functionally equivalent peptides (i.e., peptides having some degree of equivalent or similar biological activity) but differ in sequence from the

sequence shown in the Sequence Listing due to degeneracy in the genetic code, are also within the scope of the invention.

Altered polynucleotide sequences encoding polypeptides include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polynucleotide encoding a polypeptide with at least one functional characteristic of the instant polypeptides. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding the instant polypeptides, and improper or unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide sequence encoding the instant polypeptides.

Allelic variant refers to any of two or more alternative forms of a gene occupying the same chromosomal locus. Allelic variation arises naturally through mutation, and may result in phenotypic polymorphism within populations. Gene mutations can be silent (i.e., no change in the encoded polypeptide) or may encode polypeptides having altered amino acid sequence. The term allelic variant is also used herein to denote a protein encoded by an allelic variant of a gene. Splice variant refers to alternative forms of RNA transcribed from a gene. Splice variation arises naturally through use of alternative splicing sites within a transcribed RNA molecule, or less commonly between separately transcribed RNA molecules, and may result in several mRNAs transcribed from the same gene. Splice variants may encode polypeptides having altered amino acid sequence. The term splice variant is also used herein to denote a protein encoded by a splice variant of an mRNA transcribed from a gene.

Those skilled in the art would recognize that, for example, G481, SEQ ID NO: 88, represents a single transcription factor; allelic variation and alternative splicing may be expected to occur. Allelic variants of SEQ ID NO: 87 can be cloned by probing cDNA or genomic libraries from different individual organisms according to standard procedures. Allelic variants of the DNA sequence shown in SEQ ID NO: 87, including those containing silent mutations and those in which mutations result in amino acid sequence changes, are within the scope of the present invention, as are proteins which are allelic variants of SEQ ID NO: 88. cDNAs generated from alternatively spliced mRNAs, which retain the properties of the transcription factor are included within the scope of the present invention, as are polypeptides encoded by such cDNAs and mRNAs. Allelic variants and splice variants of these sequences can be cloned by probing cDNA or genomic libraries from different individual organisms or tissues according to standard procedures known in the art (see USPN 6,388,064).

Thus, in addition to the sequences set forth in the Sequence Listing, the invention also encompasses related nucleic acid molecules that include allelic or splice variants of SEQ ID NO: 2N - 1, wherein N = 1- 229, SEQ ID NO: 459-466; 468-487; 491-500; 504; 506-511; 516-520; 523-524; 527; 529; 531-533; 538-539; 541-557; 560-568; 570-586; 595-596; 598-606; 610-620; 627-634; 640-664; 670-707; 714-719; 722-735; 740-741; 743-779; 808-823; 825-834; 838-850; 855-864; 868-889; 892-902; 908-909; 914-921; 924-925; 927-932; 935-942; 944-952; 961-965; 968-986; 989-993; 995-1010; 1012-1034; 1043-1063; 1074-1080; 1091-1104; 1111-1121; 1123-1128; 1134-1138; 1142-1156; 1159-1175;

1187-1190; 1192-1199; 1202-1220; 1249-1253; 1258-1262; 1264-1269; 1271-1287; 1292-1301; 1303-1309; 1315-1323; 1328-1337; 1340-1341; 1344-1361; 1365-1377; 1379-1390; 1393-1394; 1396-1398; 1419-1432; 1434-1452; 1455-1456; 1460-1465; 1468-1491; 1499; 1502; 1505-1521; 1523-1527; 1529-1532; 1536-1539; 1542-1562; 1567-1571; 1573-1582; 1587-1592; 1595-1620; 1625-1644; 1647-1654; 1659-1669; 1671-1673; 1675-1680; 1682-1686; 1688-1700; 1706-1709; 1714-1726; 1728-1734; 1738-1742; 1744-1753; 1757-1760; 1763-1764; 1766-1768; 1770-1780; 1782-1784; 1786-1789; 1791-1804; 1806-1812; 1814-1837; 1847-1856; 1858-1862; 1864-1873; 1876-1882; 1885-1896; 1902-1910; 1913-1916; 1921-1928; 1931-1936; 1940-1941; 1944-1946, 2907-2941, 2944, 2945, 2947, 2949, or SEQ ID NO: 2N - 1, wherein N = 974- 1101, and include sequences which are complementary to any of the above nucleotide sequences. Related nucleic acid molecules also include nucleotide sequences encoding a polypeptide comprising or consisting essentially of a substitution, modification, addition and/or deletion of one or more amino acid residues compared to the polypeptide as set forth in any of SEQ ID NO: 2N, wherein N = 1- 229, SEQ ID NO: 467; 488-490; 501-503; 505; 512-515; 521-522; 525-526; 528; 530; 534-537; 540; 558-559; 569; 587-594; 597; 607-609; 621-626; 635-639; 665-669; 708-713; 720-721; 736-739; 742; 780-807; 824; 835-837; 851-854; 865-867; 890-891; 903-907; 910-913; 922-923; 926; 933-934; 943; 953-960; 966-967; 987-988; 994; 1011; 1035-1042; 1064-1073; 1081-1090; 1105-1110; 1122; 1129-1133; 1139-1141; 1157-1158; 1176-1186; 1191; 1200-1201; 1221-1248; 1254-1257; 1263; 1270; 1288-1291; 1302; 1310-1314; 1324-1327; 1338-1339; 1342-1343; 1362-1364; 1378; 1391-1392; 1395; 1399-1418; 1433; 1453-1454; 1457-1459; 1466-1467; 1492-1498; 1500-1501; 1503-1504; 1522; 1528; 1533-1535; 1540-1541; 1563-1566; 1572; 1583-1586; 1593-1594; 1621-1624; 1645-1646; 1655-1658; 1670; 1674; 1681; 1687; 1701-1705; 1710-1713; 1727; 1735-1737; 1743; 1754-1756; 1761-1762; 1765; 1769; 1781; 1785; 1790; 1805; 1813; 1838-1846; 1857; 1863; 1874-1875; 1883-1884; 1897-1901; 1911-1912; 1917-1920; 1929-1930; 1937-1939; 1942-1943; 2942 or 2943, 2945, 2947, 2949, or SEQ ID NO: 2N, wherein N = 974- 1101. Such related polypeptides may comprise, for example, additions and/or deletions of one or more N-linked or O-linked glycosylation sites, or an addition and/or a deletion of one or more cysteine residues.

For example, Table 4 illustrates, e.g., that the codons AGC, AGT, TCA, TCC, TCG, and TCT all encode the same amino acid: serine. Accordingly, at each position in the sequence where there is a codon encoding serine, any of the above trinucleotide sequences can be used without altering the encoded polypeptide.

Table 4

Amino acid			Possible Codons					
Alanine	Ala	A	GCA	GCC	GCG	GCU		
Cysteine	Cys	C	TGC	TGT				
Aspartic acid	Asp	D	GAC	GAT				
Glutamic acid	Glu	E	GAA	GAG				
Phenylalanine	Phe	F	TTC	TTT				
Glycine	Gly	G	GGA	GGC	GGG	GGT		
Histidine	His	H	CAC	CAT				
Isoleucine	Ile	I	ATA	ATC	ATT			
Lysine	Lys	K	AAA	AAG				
Leucine	Leu	L	TTA	TTG	CTA	CTC	CTG	CTT
Methionine	Met	M	ATG					
Asparagine	Asn	N	AAC	AAT				
Proline	Pro	P	CCA	CCC	CCG	CCT		
Glutamine	Gln	Q	CAA	CAG				
Arginine	Arg	R	AGA	AGG	CGA	CGC	CGG	CGT
Serine	Ser	S	AGC	AGT	TCA	TCC	TCG	TCT
Threonine	Thr	T	ACA	ACC	ACG	ACT		
Valine	Val	V	GTA	GTC	GTG	GTT		
Tryptophan	Trp	W	TGG					
Tyrosine	Tyr	Y	TAC	TAT				

Sequence alterations that do not change the amino acid sequence encoded by the polynucleotide are termed "silent" variations. With the exception of the codons ATG and TGG, encoding methionine and tryptophan, respectively, any of the possible codons for the same amino acid can be substituted by a variety of techniques, e.g., site-directed mutagenesis, available in the art. Accordingly, any and all such variations of a sequence selected from the above table are a feature of the invention.

In addition to silent variations, other conservative variations that alter one, or a few amino acids in the encoded polypeptide, can be made without altering the function of the polypeptide, these conservative variants are, likewise, a feature of the invention.

For example, substitutions, deletions and insertions introduced into the sequences provided in the Sequence Listing, are also envisioned by the invention. Such sequence modifications can be engineered into a sequence by site-directed mutagenesis (Wu (ed.) *Methods Enzymol.* (1993) vol. 217, Academic Press) or the other methods noted below. Amino acid substitutions are typically of single residues; insertions usually will be on the order of about from 1 to 10 amino acid residues; and deletions will range about from 1 to 30 residues. In preferred embodiments, deletions or insertions are made in adjacent pairs, e.g., a deletion of two residues or insertion of two residues. Substitutions, deletions, insertions or any combination thereof can be combined to arrive at a sequence. The mutations that are made in the polynucleotide encoding the transcription factor should not place the sequence out of reading frame and should not create complementary regions that could produce secondary mRNA structure. Preferably, the polypeptide encoded by the DNA performs the desired function.

Conservative substitutions are those in which at least one residue in the amino acid sequence has been removed and a different residue inserted in its place. Such substitutions generally are made in

accordance with the Table 5 when it is desired to maintain the activity of the protein. Table 5 shows amino acids which can be substituted for an amino acid in a protein and which are typically regarded as conservative substitutions.

Table 5

Residue	Conservative Substitutions
Ala	Ser
Arg	Lys
Asn	Gln; His
Asp	Glu
Gln	Asn
Cys	Ser
Glu	Asp
Gly	Pro
His	Asn; Gln
Ile	Leu, Val
Leu	Ile; Val
Lys	Arg; Gln
Met	Leu; Ile
Phe	Met; Leu; Tyr
Ser	Thr; Gly
Thr	Ser; Val
Trp	Tyr
Tyr	Trp; Phe
Val	Ile; Leu

Similar substitutions are those in which at least one residue in the amino acid sequence has been removed and a different residue inserted in its place. Such substitutions generally are made in accordance with the Table 6 when it is desired to maintain the activity of the protein. Table 6 shows amino acids which can be substituted for an amino acid in a protein and which are typically regarded as structural and functional substitutions. For example, a residue in column 1 of Table 6 may be substituted with a residue in column 2; in addition, a residue in column 2 of Table 6 may be substituted with the residue of column 1.

Table 6

Residue	Similar Substitutions
Ala	Ser; Thr; Gly; Val; Leu; Ile
Arg	Lys; His; Gly
Asn	Gln; His; Gly; Ser; Thr
Asp	Glu, Ser; Thr
Gln	Asn; Ala
Cys	Ser; Gly
Glu	Asp
Gly	Pro; Arg
His	Asn; Gln; Tyr; Phe; Lys; Arg
Ile	Ala; Leu; Val; Gly; Met
Leu	Ala; Ile; Val; Gly; Met
Lys	Arg; His; Gln; Gly; Pro
Met	Leu; Ile; Phe
Phe	Met; Leu; Tyr; Trp; His; Val; Ala
Ser	Thr; Gly; Asp; Ala; Val; Ile; His
Thr	Ser; Val; Ala; Gly
Trp	Tyr; Phe; His
Tyr	Trp; Phe; His
Val	Ala; Ile; Leu; Gly; Thr; Ser; Glu

Substitutions that are less conservative than those in Table 5 can be selected by picking residues that differ more significantly in their effect on maintaining (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a sheet or helical conformation, (b) the charge or hydrophobicity of the molecule at the target site, or (c) the bulk of the side chain. The substitutions which in general are expected to produce the greatest changes in protein properties will be those in which (a) a hydrophilic residue, e.g., seryl or threonyl, is substituted for (or by) a hydrophobic residue, e.g., leucyl, isoleucyl, phenylalanyl, valyl or alanyl; (b) a cysteine or proline is substituted for (or by) any other residue; (c) a residue having an electropositive side chain, e.g., lysyl, arginyl, or histidyl, is substituted for (or by) an electronegative residue, e.g., glutamyl or aspartyl; or (d) a residue having a bulky side chain, e.g., phenylalanine, is substituted for (or by) one not having a side chain, e.g., glycine.

Further Modifying Sequences of the Invention – Mutation/Forced Evolution

In addition to generating silent or conservative substitutions as noted, above, the present invention optionally includes methods of modifying the sequences of the Sequence Listing. In the methods, nucleic acid or protein modification methods are used to alter the given sequences to produce new sequences and/or to chemically or enzymatically modify given sequences to change the properties of the nucleic acids or proteins.

Thus, in one embodiment, given nucleic acid sequences are modified, e.g., according to standard mutagenesis or artificial evolution methods to produce modified sequences. The modified sequences may be created using purified natural polynucleotides isolated from any organism or may be synthesized from purified compositions and chemicals using chemical means well know to those of skill in the art. For example, Ausubel, *supra*, provides additional details on mutagenesis methods. Artificial forced evolution methods are described, for example, by Stemmer (1994) *Nature* 370: 389-391, Stemmer (1994) *Proc. Natl. Acad. Sci.* 91: 10747-10751, and U.S. Patents 5,811,238, 5,837,500, and 6,242,568. Methods for engineering synthetic transcription factors and other polypeptides are described, for example, by Zhang et al. (2000) *J. Biol. Chem.* 275: 33850-33860, Liu et al. (2001) *J. Biol. Chem.* 276: 11323-11334, and Isalan et al. (2001) *Nature Biotechnol.* 19: 656-660. Many other mutation and evolution methods are also available and expected to be within the skill of the practitioner.

Similarly, chemical or enzymatic alteration of expressed nucleic acids and polypeptides can be performed by standard methods. For example, sequence can be modified by addition of lipids, sugars, peptides, organic or inorganic compounds, by the inclusion of modified nucleotides or amino acids, or the like. For example, protein modification techniques are illustrated in Ausubel, *supra*. Further details on chemical and enzymatic modifications can be found herein. These modification methods can be used to modify any given sequence, or to modify any sequence produced by the various mutation and artificial evolution modification methods noted herein.

Accordingly, the invention provides for modification of any given nucleic acid by mutation, evolution, chemical or enzymatic modification, or other available methods, as well as for the products produced by practicing such methods, e.g., using the sequences herein as a starting substrate for the various modification approaches.

For example, optimized coding sequence containing codons preferred by a particular prokaryotic or eukaryotic host can be used e.g., to increase the rate of translation or to produce recombinant RNA transcripts having desirable properties, such as a longer half-life, as compared with transcripts produced using a non-optimized sequence. Translation stop codons can also be modified to reflect host preference. For example, preferred stop codons for *Saccharomyces cerevisiae* and mammals are TAA and TGA, respectively. The preferred stop codon for monocotyledonous plants is TGA, whereas insects and *E. coli* prefer to use TAA as the stop codon.

The polynucleotide sequences of the present invention can also be engineered in order to alter a coding sequence for a variety of reasons, including but not limited to, alterations which modify the

sequence to facilitate cloning, processing and/or expression of the gene product. For example, alterations are optionally introduced using techniques which are well known in the art, e.g., site-directed mutagenesis, to insert new restriction sites, to alter glycosylation patterns, to change codon preference, to introduce splice sites, etc.

Furthermore, a fragment or domain derived from any of the polypeptides of the invention can be combined with domains derived from other transcription factors or synthetic domains to modify the biological activity of a transcription factor. For instance, a DNA-binding domain derived from a transcription factor of the invention can be combined with the activation domain of another transcription factor or with a synthetic activation domain. A transcription activation domain assists in initiating transcription from a DNA-binding site. Examples include the transcription activation region of VP16 or GAL4 (Moore et al. (1998) *Proc. Natl. Acad. Sci.* 95: 376-381; Aoyama et al. (1995) *Plant Cell* 7: 1773-1785), peptides derived from bacterial sequences (Ma and Ptashne (1987) *Cell* 51: 113-119) and synthetic peptides (Giniger and Ptashne (1987) *Nature* 330: 670-672).

EXPRESSION AND MODIFICATION OF POLYPEPTIDES

Typically, polynucleotide sequences of the invention are incorporated into recombinant DNA (or RNA) molecules that direct expression of polypeptides of the invention in appropriate host cells, transgenic plants, in vitro translation systems, or the like. Due to the inherent degeneracy of the genetic code, nucleic acid sequences which encode substantially the same or a functionally equivalent amino acid sequence can be substituted for any listed sequence to provide for cloning and expressing the relevant homolog.

The transgenic plants of the present invention comprising recombinant polynucleotide sequences are generally derived from parental plants, which may themselves be non-transformed (or non-transgenic) plants. These transgenic plants may either have a transcription factor gene "knocked out" (for example, with a genomic insertion by homologous recombination, an antisense or ribozyme construct) or expressed to a normal or wild-type extent. However, overexpressing transgenic "progeny" plants will exhibit greater mRNA levels, wherein the mRNA encodes a transcription factor, that is, a DNA-binding protein that is capable of binding to a DNA regulatory sequence and inducing transcription, and preferably, expression of a plant trait gene. Preferably, the mRNA expression level will be at least three-fold greater than that of the parental plant, or more preferably at least ten-fold greater mRNA levels compared to said parental plant, and most preferably at least fifty-fold greater compared to said parental plant.

Vectors, Promoters, and Expression Systems

The present invention includes recombinant constructs comprising one or more of the nucleic acid sequences herein. The constructs typically comprise a vector, such as a plasmid, a cosmid, a phage, a virus (e.g., a plant virus), a bacterial artificial chromosome (BAC), a yeast artificial chromosome (YAC),

or the like, into which a nucleic acid sequence of the invention has been inserted, in a forward or reverse orientation. In a preferred aspect of this embodiment, the construct further comprises regulatory sequences, including, for example, a promoter, operably linked to the sequence. Large numbers of suitable vectors and promoters are known to those of skill in the art, and are commercially available.

General texts that describe molecular biological techniques useful herein, including the use and production of vectors, promoters and many other relevant topics, include Berger, Sambrook, *supra* and Ausubel, *supra*. Any of the identified sequences can be incorporated into a cassette or vector, e.g., for expression in plants. A number of expression vectors suitable for stable transformation of plant cells or for the establishment of transgenic plants have been described including those described in Weissbach and Weissbach (1989) *Methods for Plant Molecular Biology*, Academic Press, and Gelvin et al. (1990) *Plant Molecular Biology Manual*, Kluwer Academic Publishers. Specific examples include those derived from a Ti plasmid of *Agrobacterium tumefaciens*, as well as those disclosed by Herrera-Estrella et al. (1983) *Nature* 303: 209, Bevan (1984) *Nucleic Acids Res.* 12: 8711-8721, Klee (1985) *Bio/Technology* 3: 637-642, for dicotyledonous plants.

Alternatively, non-Ti vectors can be used to transfer the DNA into monocotyledonous plants and cells by using free DNA delivery techniques. Such methods can involve, for example, the use of liposomes, electroporation, microprojectile bombardment, silicon carbide whiskers, and viruses. By using these methods transgenic plants such as wheat, rice (Christou (1991) *Bio/Technology* 9: 957-962) and corn (Gordon-Kamm (1990) *Plant Cell* 2: 603-618) can be produced. An immature embryo can also be a good target tissue for monocots for direct DNA delivery techniques by using the particle gun (Weeks et al. (1993) *Plant Physiol.* 102: 1077-1084; Vasil (1993) *Bio/Technology* 10: 667-674; Wan and Lemeaux (1994) *Plant Physiol.* 104: 37-48, and for *Agrobacterium*-mediated DNA transfer (Ishida et al. (1996) *Nature Biotechnol.* 14: 745-750).

Typically, plant transformation vectors include one or more cloned plant coding sequence (genomic or cDNA) under the transcriptional control of 5' and 3' regulatory sequences and a dominant selectable marker. Such plant transformation vectors typically also contain a promoter (e.g., a regulatory region controlling inducible or constitutive, environmentally-or developmentally-regulated, or cell- or tissue-specific expression), a transcription initiation start site, an RNA processing signal (such as intron splice sites), a transcription termination site, and/or a polyadenylation signal.

A potential utility for the transcription factor polynucleotides disclosed herein is the isolation of promoter elements from these genes that can be used to program expression in plants of any genes. Each transcription factor gene disclosed herein is expressed in a unique fashion, as determined by promoter elements located upstream of the start of translation, and additionally within an intron of the transcription factor gene or downstream of the termination codon of the gene. As is well known in the art, for a significant portion of genes, the promoter sequences are located entirely in the region directly upstream of the start of translation. In such cases, typically the promoter sequences are located within 2.0 kb of the

start of translation, or within 1.5 kb of the start of translation, frequently within 1.0 kb of the start of translation, and sometimes within 0.5 kb of the start of translation.

The promoter sequences can be isolated according to methods known to one skilled in the art.

Examples of constitutive plant promoters which can be useful for expressing the TF sequence include: the cauliflower mosaic virus (CaMV) 35S promoter, which confers constitutive, high-level expression in most plant tissues (*see,* e.g., Odell et al. (1985) *Nature* 313: 810-812); the nopaline synthase promoter (An et al. (1988) *Plant Physiol.* 88: 547-552); and the octopine synthase promoter (Fromm et al. (1989) *Plant Cell* 1: 977-984).

A variety of plant gene promoters that regulate gene expression in response to environmental, hormonal, chemical, developmental signals, and in a tissue-active manner can be used for expression of a TF sequence in plants. Choice of a promoter is based largely on the phenotype of interest and is determined by such factors as tissue (e.g., seed, fruit, root, pollen, vascular tissue, flower, carpel, etc.), inducibility (e.g., in response to wounding, heat, cold, drought, light, pathogens, etc.), timing, developmental stage, and the like. Numerous known promoters have been characterized and can favorably be employed to promote expression of a polynucleotide of the invention in a transgenic plant or cell of interest. For example, tissue specific promoters include: seed-specific promoters (such as the napin, phaseolin or DC3 promoter described in US Pat. No. 5,773,697), fruit-specific promoters that are active during fruit ripening (such as the dru 1 promoter (US Pat. No. 5,783,393), or the 2A11 promoter (US Pat. No. 4,943,674) and the tomato polygalacturonase promoter (Bird et al. (1988) *Plant Mol. Biol.* 11: 651-662), root-specific promoters, such as those disclosed in US Patent Nos. 5,618,988, 5,837,848 and 5,905,186, pollen-active promoters such as PTA29, PTA26 and PTA13 (US Pat. No. 5,792,929), promoters active in vascular tissue (Ringli and Keller (1998) *Plant Mol. Biol.* 37: 977-988), flower-specific (Kaiser et al. (1995) *Plant Mol. Biol.* 28: 231-243), pollen (Baerson et al. (1994) *Plant Mol. Biol.* 26: 1947-1959), carpels (Ohl et al. (1990) *Plant Cell* 2: 837-848), pollen and ovules (Baerson et al. (1993) *Plant Mol. Biol.* 22: 255-267), auxin-inducible promoters (such as that described in van der Kop et al. (1999) *Plant Mol. Biol.* 39: 979-990 or Baumann et al. (1999) *Plant Cell* 11: 323-334), cytokinin-inducible promoter (Guevara-Garcia (1998) *Plant Mol. Biol.* 38: 743-753), promoters responsive to gibberellin (Shi et al. (1998) *Plant Mol. Biol.* 38: 1053-1060, Willmott et al. (1998) 38: 817-825) and the like. Additional promoters are those that elicit expression in response to heat (Ainley et al. (1993) *Plant Mol. Biol.* 22: 13-23), light (e.g., the pea rbcS-3A promoter, Kuhlemeier et al. (1989) *Plant Cell* 1: 471-478, and the maize rbcS promoter, Schaffner and Sheen (1991) *Plant Cell* 3: 997-1012); wounding (e.g., *wunI*, Siebertz et al. (1989) *Plant Cell* 1: 961-968); pathogens (such as the PR-1 promoter described in Buchel et al. (1999) *Plant Mol. Biol.* 40: 387-396, and the PDF1.2 promoter described in Manners et al. (1998) *Plant Mol. Biol.* 38: 1071-1080), and chemicals such as methyl jasmonate or salicylic acid (Gatz (1997) *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 48: 89-108). In addition, the timing of the expression can be controlled by using promoters such as those acting at senescence (Gan and Amasino (1995) *Science* 270: 1986-1988); or late seed development (Odell et al. (1994) *Plant Physiol.* 106: 447-458).

Plant expression vectors can also include RNA processing signals that can be positioned within, upstream or downstream of the coding sequence. In addition, the expression vectors can include additional regulatory sequences from the 3'-untranslated region of plant genes, e.g., a 3' terminator region to increase mRNA stability of the mRNA, such as the PI-II terminator region of potato or the octopine or nopaline synthase 3' terminator regions.

Additional Expression Elements

Specific initiation signals can aid in efficient translation of coding sequences. These signals can include, e.g., the ATG initiation codon and adjacent sequences. In cases where a coding sequence, its initiation codon and upstream sequences are inserted into the appropriate expression vector, no additional translational control signals may be needed. However, in cases where only coding sequence (e.g., a mature protein coding sequence), or a portion thereof, is inserted, exogenous transcriptional control signals including the ATG initiation codon can be separately provided. The initiation codon is provided in the correct reading frame to facilitate transcription. Exogenous transcriptional elements and initiation codons can be of various origins, both natural and synthetic. The efficiency of expression can be enhanced by the inclusion of enhancers appropriate to the cell system in use.

Expression Hosts

The present invention also relates to host cells which are transduced with vectors of the invention, and the production of polypeptides of the invention (including fragments thereof) by recombinant techniques. Host cells are genetically engineered (i.e., nucleic acids are introduced, e.g., transduced, transformed or transfected) with the vectors of this invention, which may be, for example, a cloning vector or an expression vector comprising the relevant nucleic acids herein. The vector is optionally a plasmid, a viral particle, a phage, a naked nucleic acid, *etc.* The engineered host cells can be cultured in conventional nutrient media modified as appropriate for activating promoters, selecting transformants, or amplifying the relevant gene. The culture conditions, such as temperature, pH and the like, are those previously used with the host cell selected for expression, and will be apparent to those skilled in the art and in the references cited herein, including, Sambrook, *supra* and Ausubel, *supra*.

The host cell can be a eukaryotic cell, such as a yeast cell, or a plant cell, or the host cell can be a prokaryotic cell, such as a bacterial cell. Plant protoplasts are also suitable for some applications. For example, the DNA fragments are introduced into plant tissues, cultured plant cells or plant protoplasts by standard methods including electroporation (Fromm et al. (1985) *Proc. Natl. Acad. Sci.* 82: 5824-5828, infection by viral vectors such as cauliflower mosaic virus (CaMV) (Hohn et al. (1982) *Molecular Biology of Plant* Tumors Academic Press, New York, NY, pp. 549-560; US 4,407,956), high velocity ballistic penetration by small particles with the nucleic acid either within the matrix of small beads or particles, or on the surface (Klein et al. (1987) *Nature* 327: 70-73), use of pollen as vector (WO 85/01856), or use of *Agrobacterium tumefaciens* or *A. rhizogenes* carrying a T-DNA plasmid in which

DNA fragments are cloned. The T-DNA plasmid is transmitted to plant cells upon infection by *Agrobacterium tumefaciens*, and a portion is stably integrated into the plant genome (Horsch et al. (1984) *Science* 233: 496-498; Fraley et al. (1983) *Proc. Natl. Acad. Sci.* 80: 4803-4807).

The cell can include a nucleic acid of the invention that encodes a polypeptide, wherein the cell expresses a polypeptide of the invention. The cell can also include vector sequences, or the like. Furthermore, cells and transgenic plants that include any polypeptide or nucleic acid above or throughout this specification, e.g., produced by transduction of a vector of the invention, are an additional feature of the invention.

For long-term, high-yield production of recombinant proteins, stable expression can be used. Host cells transformed with a nucleotide sequence encoding a polypeptide of the invention are optionally cultured under conditions suitable for the expression and recovery of the encoded protein from cell culture. The protein or fragment thereof produced by a recombinant cell may be secreted, membrane-bound, or contained intracellularly, depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides encoding mature proteins of the invention can be designed with signal sequences which direct secretion of the mature polypeptides through a prokaryotic or eukaryotic cell membrane.

Modified Amino Acid Residues

Polypeptides of the invention may contain one or more modified amino acid residues. The presence of modified amino acids may be advantageous in, for example, increasing polypeptide half-life, reducing polypeptide antigenicity or toxicity, increasing polypeptide storage stability, or the like. Amino acid residue(s) are modified, for example, co-translationally or post-translationally during recombinant production or modified by synthetic or chemical means.

Non-limiting examples of a modified amino acid residue include incorporation or other use of acetylated amino acids, glycosylated amino acids, sulfated amino acids, prenylated (e.g., farnesylated, geranylgeranylated) amino acids, PEG modified (e.g., "PEGylated") amino acids, biotinylated amino acids, carboxylated amino acids, phosphorylated amino acids, etc. References adequate to guide one of skill in the modification of amino acid residues are replete throughout the literature.

The modified amino acid residues may prevent or increase affinity of the polypeptide for another molecule, including, but not limited to, polynucleotide, proteins, carbohydrates, lipids and lipid derivatives, and other organic or synthetic compounds.

Identification of Additional Factors

A transcription factor provided by the present invention can also be used to identify additional endogenous or exogenous molecules that can affect a phentoype or trait of interest. On the one hand, such molecules include organic (small or large molecules) and/or inorganic compounds that affect expression of (i.e., regulate) a particular transcription factor. Alternatively, such molecules include endogenous

molecules that are acted upon either at a transcriptional level by a transcription factor of the invention to modify a phenotype as desired. For example, the transcription factors can be employed to identify one or more downstream genes that are subject to a regulatory effect of the transcription factor. In one approach, a transcription factor or transcription factor homolog of the invention is expressed in a host cell, e.g., a transgenic plant cell, tissue or explant, and expression products, either RNA or protein, of likely or random targets are monitored, e.g., by hybridization to a microarray of nucleic acid probes corresponding to genes expressed in a tissue or cell type of interest, by two-dimensional gel electrophoresis of protein products, or by any other method known in the art for assessing expression of gene products at the level of RNA or protein. Alternatively, a transcription factor of the invention can be used to identify promoter sequences (such as binding sites on DNA sequences) involved in the regulation of a downstream target. After identifying a promoter sequence, interactions between the transcription factor and the promoter sequence can be modified by changing specific nucleotides in the promoter sequence or specific amino acids in the transcription factor that interact with the promoter sequence to alter a plant trait. Typically, transcription factor DNA- binding sites are identified by gel shift assays. After identifying the promoter regions, the promoter region sequences can be employed in double-stranded DNA arrays to identify molecules that affect the interactions of the transcription factors with their promoters (Bulyk et al. (1999) *Nature Biotechnol.* 17: 573-577).

The identified transcription factors are also useful to identify proteins that modify the activity of the transcription factor. Such modification can occur by covalent modification, such as by phosphorylation, or by protein-protein (homo or-heteropolymer) interactions. Any method suitable for detecting protein-protein interactions can be employed. Among the methods that can be employed are co-immunoprecipitation, cross-linking and co-purification through gradients or chromatographic columns, and the two-hybrid yeast system.

The two-hybrid system detects protein interactions in vivo and is described in Chien et al. (1991) *Proc. Natl. Acad. Sci.* 88: 9578-9582, and is commercially available from Clontech (Palo Alto, Calif.). In such a system, plasmids are constructed that encode two hybrid proteins: one consists of the DNA-binding domain of a transcription activator protein fused to the TF polypeptide and the other consists of the transcription activator protein's activation domain fused to an unknown protein that is encoded by a cDNA that has been recombined into the plasmid as part of a cDNA library. The DNA-binding domain fusion plasmid and the cDNA library are transformed into a strain of the yeast *Saccharomyces cerevisiae* that contains a reporter gene (e.g., lacZ) whose regulatory region contains the transcription activator's binding site. Either hybrid protein alone cannot activate transcription of the reporter gene. Interaction of the two hybrid proteins reconstitutes the functional activator protein and results in expression of the reporter gene, which is detected by an assay for the reporter gene product. Then, the library plasmids responsible for reporter gene expression are isolated and sequenced to identify the proteins encoded by the library plasmids. After identifying proteins that interact with the transcription factors, assays for compounds that interfere with the TF protein-protein interactions can be preformed.

Identification of Modulators

In addition to the intracellular molecules described above, extracellular molecules that alter activity or expression of a transcription factor, either directly or indirectly, can be identified. For example, the methods can entail first placing a candidate molecule in contact with a plant or plant cell. The molecule can be introduced by topical administration, such as spraying or soaking of a plant, or incubating a plant in a solution containing the molecule, and then the molecule's effect on the expression or activity of the TF polypeptide or the expression of the polynucleotide monitored. Changes in the expression of the TF polypeptide can be monitored by use of polyclonal or monoclonal antibodies, gel electrophoresis or the like. Changes in the expression of the corresponding polynucleotide sequence can be detected by use of microarrays, Northerns, quantitative PCR, or any other technique for monitoring changes in mRNA expression. These techniques are exemplified in Ausubel et al. (eds.) *Current Protocols in Molecular Biology*, John Wiley & Sons (1998, and supplements through 2001).Changes in the activity of the transcription factor can be monitored, directly or indirectly, by assaying the function of the transcription factor, for example, by measuring the expression of promoters known to be controlled by the transcription factor (using promoter-reporter constructs), measuring the levels of transcripts using microarrays, Northern blots, quantitative PCR, etc. Such changes in the expression levels can be correlated with modified plant traits and thus identified molecules can be useful for soaking or spraying on fruit, vegetable and grain crops to modify traits in plants.

Essentially any available composition can be tested for modulatory activity of expression or activity of any nucleic acid or polypeptide herein. Thus, available libraries of compounds such as chemicals, polypeptides, nucleic acids and the like can be tested for modulatory activity. Often, potential modulator compounds can be dissolved in aqueous or organic (e.g., DMSO-based) solutions for easy delivery to the cell or plant of interest in which the activity of the modulator is to be tested. Optionally, the assays are designed to screen large modulator composition libraries by automating the assay steps and providing compounds from any convenient source to assays, which are typically run in parallel (e.g., in microtiter formats on microplates in robotic assays).

In one embodiment, high throughput screening methods involve providing a combinatorial library containing a large number of potential compounds (potential modulator compounds). Such "combinatorial chemical libraries" are then screened in one or more assays, as described herein, to identify those library members (particular chemical species or subclasses) that display a desired characteristic activity. The compounds thus identified can serve as target compounds.

A combinatorial chemical library can be, e.g., a collection of diverse chemical compounds generated by chemical synthesis or biological synthesis. For example, a combinatorial chemical library such as a polypeptide library is formed by combining a set of chemical building blocks (e.g., in one example, amino acids) in every possible way for a given compound length (i.e., the number of amino acids in a polypeptide compound of a set length). Exemplary libraries include peptide libraries, nucleic

acid libraries, antibody libraries (see, e.g., Vaughn et al. (1996) *Nature Biotechnol.* 14: 309-314 and PCT/US96/10287), carbohydrate libraries (see, e.g., Liang et al. *Science* (1996) 274: 1520-1522 and U.S. Patent 5,593,853), peptide nucleic acid libraries (see, e.g., U.S. Patent 5,539,083), and small organic molecule libraries (see, e.g., benzodiazepines, in Baum *Chem. & Engineering News* Jan 18, 1993, page 33; isoprenoids, U.S. Patent 5,569,588; thiazolidinones and metathiazanones, U.S. Patent 5,549,974; pyrrolidines, U.S. Patents 5,525,735 and 5,519,134; morpholino compounds, U.S. Patent 5,506,337) and the like.

Preparation and screening of combinatorial or other libraries is well known to those of skill in the art. Such combinatorial chemical libraries include, but are not limited to, peptide libraries (see, e.g., U.S. Patent 5,010,175; Furka, (1991) *Int. J. Pept. Prot. Res.* 37: 487-493; and Houghton et al. (1991) *Nature* 354: 84-88). Other chemistries for generating chemical diversity libraries can also be used.

In addition, as noted, compound screening equipment for high-throughput screening is generally available, e.g., using any of a number of well known robotic systems that have also been developed for solution phase chemistries useful in assay systems. These systems include automated workstations including an automated synthesis apparatus and robotic systems utilizing robotic arms. Any of the above devices are suitable for use with the present invention, e.g., for high-throughput screening of potential modulators. The nature and implementation of modifications to these devices (if any) so that they can operate as discussed herein will be apparent to persons skilled in the relevant art.

Indeed, entire high-throughput screening systems are commercially available. These systems typically automate entire procedures including all sample and reagent pipetting, liquid dispensing, timed incubations, and final readings of the microplate in detector(s) appropriate for the assay. These configurable systems provide high throughput and rapid start up as well as a high degree of flexibility and customization. Similarly, microfluidic implementations of screening are also commercially available.

The manufacturers of such systems provide detailed protocols the various high throughput. Thus, for example, Zymark Corp. provides technical bulletins describing screening systems for detecting the modulation of gene transcription, ligand binding, and the like. The integrated systems herein, in addition to providing for sequence alignment and, optionally, synthesis of relevant nucleic acids, can include such screening apparatus to identify modulators that have an effect on one or more polynucleotides or polypeptides according to the present invention.

In some assays it is desirable to have positive controls to ensure that the components of the assays are working properly. At least two types of positive controls are appropriate. That is, known transcriptional activators or inhibitors can be incubated with cells or plants, for example, in one sample of the assay, and the resulting increase/decrease in transcription can be detected by measuring the resulting increase in RNA levels and/or protein expression, for example, according to the methods herein. It will be appreciated that modulators can also be combined with transcriptional activators or inhibitors to find modulators that inhibit transcriptional activation or transcriptional repression. Either expression of the

nucleic acids and proteins herein or any additional nucleic acids or proteins activated by the nucleic acids or proteins herein, or both, can be monitored.

In an embodiment, the invention provides a method for identifying compositions that modulate the activity or expression of a polynucleotide or polypeptide of the invention. For example, a test compound, whether a small or large molecule, is placed in contact with a cell, plant (or plant tissue or explant), or composition comprising the polynucleotide or polypeptide of interest and a resulting effect on the cell, plant, (or tissue or explant) or composition is evaluated by monitoring, either directly or indirectly, one or more of: expression level of the polynucleotide or polypeptide, activity (or modulation of the activity) of the polynucleotide or polypeptide. In some cases, an alteration in a plant phenotype can be detected following contact of a plant (or plant cell, or tissue or explant) with the putative modulator, e.g., by modulation of expression or activity of a polynucleotide or polypeptide of the invention. Modulation of expression or activity of a polynucleotide or polypeptide of the invention may also be caused by molecular elements in a signal transduction second messenger pathway and such modulation can affect similar elements in the same or another signal transduction second messenger pathway.

Subsequences

Also contemplated are uses of polynucleotides, also referred to herein as oligonucleotides, typically having at least 12 bases, preferably at least 15, more preferably at least 20, 30, or 50 bases, which hybridize under at least highly stringent (or ultra-high stringent or ultra-ultra-high stringent conditions) conditions to a polynucleotide sequence described above. The polynucleotides may be used as probes, primers, sense and antisense agents, and the like, according to methods as noted *supra*.

Subsequences of the polynucleotides of the invention, including polynucleotide fragments and oligonucleotides are useful as nucleic acid probes and primers. An oligonucleotide suitable for use as a probe or primer is at least about 15 nucleotides in length, more often at least about 18 nucleotides, often at least about 21 nucleotides, frequently at least about 30 nucleotides, or about 40 nucleotides, or more in length. A nucleic acid probe is useful in hybridization protocols, e.g., to identify additional polypeptide homologs of the invention, including protocols for microarray experiments. Primers can be annealed to a complementary target DNA strand by nucleic acid hybridization to form a hybrid between the primer and the target DNA strand, and then extended along the target DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification of a nucleic acid sequence, e.g., by the polymerase chain reaction (PCR) or other nucleic-acid amplification methods. See Sambrook, *supra*, and Ausubel, *supra*.

In addition, the invention includes an isolated or recombinant polypeptide including a subsequence of at least about 15 contiguous amino acids encoded by the recombinant or isolated polynucleotides of the invention. For example, such polypeptides, or domains or fragments thereof, can be used as immunogens, e.g., to produce antibodies specific for the polypeptide sequence, or as probes for detecting a sequence of interest. A subsequence can range in size from about 15 amino acids in length up to and including the full length of the polypeptide.

To be encompassed by the present invention, an expressed polypeptide which comprises such a polypeptide subsequence performs at least one biological function of the intact polypeptide in substantially the same manner, or to a similar extent, as does the intact polypeptide. For example, a polypeptide fragment can comprise a recognizable structural motif or functional domain such as a DNA binding domain that activates transcription, e.g., by binding to a specific DNA promoter region an activation domain, or a domain for protein-protein interactions.

Production of Transgenic Plants

Modification of Traits

The polynucleotides of the invention are favorably employed to produce transgenic plants with various traits, or characteristics, that have been modified in a desirable manner, e.g., to improve the seed characteristics of a plant. For example, alteration of expression levels or patterns (e.g., spatial or temporal expression patterns) of one or more of the transcription factors (or transcription factor homologs) of the invention, as compared with the levels of the same protein found in a wild-type plant, can be used to modify a plant's traits. An illustrative example of trait modification, improved characteristics, by altering expression levels of a particular transcription factor is described further in the Examples and the Sequence Listing.

Arabidopsis as a model system

Arabidopsis thaliana is the object of rapidly growing attention as a model for genetics and metabolism in plants. *Arabidopsis* has a small genome, and well-documented studies are available. It is easy to grow in large numbers and mutants defining important genetically controlled mechanisms are either available, or can readily be obtained. Various methods to introduce and express isolated homologous genes are available (see Koncz et al. eds., et al. Methods in *Arabidopsis* Researeh (1992) et al. World Scientific, New Jersey, NJ, in "Preface"). Because of its small size, short life cycle, obligate autogamy and high fertility, *Arabidopsis* is also a choice organism for the isolation of mutants and studies in morphogenetic and development pathways, and control of these pathways by transcription factors (Koncz *supra*, p. 72). A number of studies introducing transcription factors into *A. thaliana* have demonstrated the utility of this plant for understanding the mechanisms of gene regulation and trait alteration in plants. (See, for example, Koncz *supra*, and U.S. Patent Number 6,417,428).

Arabidopsis genes in transgenic plants.

Expression of genes which encode transcription factors modify expression of endogenous genes, polynucleotides, and proteins are well known in the art. In addition, transgenic plants comprising isolated polynucleotides encoding transcription factors may also modify expression of endogenous genes, polynucleotides, and proteins. Examples include Peng et al. (1997) et al. *Genes and Development* 11: 3194-3205, and Peng et al. (1999) *Nature* 400: 256-261. In addition, many others have demonstrated that

an *Arabidopsis* transcription factor expressed in an exogenous plant species elicits the same or very similar phenotypic response. See, for example, Fu et al. (2001) *Plant Cell* 13: 1791-1802; Nandi et al. (2000) *Curr. Biol.* 10: 215-218; Coupland (1995) *Nature* 377: 482-483; and Weigel and Nilsson (1995) *Nature* 377: 482-500.

Homologous genes introduced into transgenic plants.

Homologous genes that may be derived from any plant, or from any source whether natural, synthetic, semi-synthetic or recombinant, and that share significant sequence identity or similarity to those provided by the present invention, may be introduced into plants, for example, crop plants, to confer desirable or improved traits. Consequently, transgenic plants may be produced that comprise a recombinant expression vector or cassette with a promoter operably linked to one or more sequences homologous to presently disclosed sequences. The promoter may be, for example, a plant or viral promoter.

The invention thus provides for methods for preparing transgenic plants, and for modifying plant traits. These methods include introducing into a plant a recombinant expression vector or cassette comprising a functional promoter operably linked to one or more sequences homologous to presently disclosed sequences. Plants and kits for producing these plants that result from the application of these methods are also encompassed by the present invention.

Transcription factors of interest for the modification of plant traits

Currently, the existence of a series of maturity groups for different latitudes represents a major barrier to the introduction of new valuable traits. Any trait (e.g. disease resistance) has to be bred into each of the different maturity groups separately, a laborious and costly exercise. The availability of single strain, which could be grown at any latitude, would therefore greatly increase the potential for introducing new traits to crop species such as soybean and cotton.

For many of the specific effects, traits and utilities listed in Table 7 and Table 9 that may be conferred to plants, one or more transcription factor genes may be used to increase or decrease, advance or delay, or improve or prove deleterious to a given trait. Overexpressing or suppressing one or more genes can impart significant differences in production of plant products, such as different fatty acid ratios. For example, overexpression of G720 caused a plant to become more freezing tolerant, but knocking out the same transcription factor imparted greater susceptibility to freezing. Thus, suppressing a gene that causes a plant to be more sensitive to cold may improve a plant's tolerance of cold.

More than one transcription factor gene may be introduced into a plant, either by transforming the plant with one or more vectors comprising two or more transcription factors, or by selective breeding of plants to yield hybrid crosses that comprise more than one introduced transcription factor. Transgenic plants may be crossed with another plant or selfed or to produce seed; which may be used to generate progeny plants having increased tolerance to abiotic stress ("selfing" refers to self-pollinating, or using

pollen from one plant to fertilize the same plant or another plant in the same line, whereas "crossing" generally refers to cross pollination with plant from a different line, such as a non-transformed or wild-type plant, or another transformed plant from a different transgenic line of plants). Crossing provides the advantage of being able to produce new varieties. The resulting seed may then be used to grow a progeny plant that is transgenic and has increased tolerance to abiotic stress. Generally, the progeny plants will express mRNA that encodes a DNA-binding protein having a conserved domain (e.g., an AP2, MYB-related or CCAAT-box binding domain) that binds to a DNA molecule, regulates its expression, and induces the expression of genes and polypeptides that confer to the plant the desirable trait (e.g., abiotic stress tolerance). In these progeny plants, the mRNA may be expressed at a level greater than in a non-transformed plant that does not overexpress the DNA-binding protein.

A listing of specific effects and utilities that the presently disclosed transcription factor genes have on plants, as determined by direct observation and assay analysis, is provided in Table 7. Table 7 shows the polynucleotides identified by SEQ ID NO; GID; and whether the polynucleotide was tested in a transgenic assay. The first column shows the polynucleotide SEQ ID NO; the second column shows the GID; the third column shows whether the gene was overexpressed (OE) or knocked out (KO) in plant studies; the fourth column shows the trait(s) resulting from the knock out or overexpression of the polynucleotide in the transgenic plant; and the fifth column ("Observations"), includes specific experimental observations made when expression of the polynucleotide of the first column was altered.

Table 7. Traits categories and effects of transcription factor genes that are overexpressed (OE) or knocked out (KO)

Polynucleotide SEQ ID NO:	GID	OE/KO	Trait Alterations	Observations
1	G8	OE	Altered flowering time Growth regulation ; nutrient uptake	Late flowering Altered C/N sensing
3	G19	OE	Increased tolerance to disease	Increased tolerance to *Erysiphe*; repressed by methyl jasmonate and induced by 1-aminocyclopropane 1-carboxylic acid (ACC)
5	G22	OE	Increased tolerance to abiotic stress	Increased tolerance to high salt
7	G24	OE	Altered necrosis Growth regulation ; nutrient uptake	Reduced size and necrotic patches Altered C/N sensing
2217	G27	OE	Growth regulation ; nutrient uptake	Altered C/N sensing
9	G28	OE	Increased tolerance to disease	Increased tolerance to *Botrytis* Increased tolerance to *Sclerotinia* Increased resistance to *Erysiphe*
11	G47	OE	Altered stem morphology Increased tolerance to abiotic stress Altered flowering time	Altered structure of vascular tissues Better root growth under osmotic stress Late flowering Altered architecture and inflorescence

			Altered architecture Increased tolerance to abiotic and osmotic stress	development Reduced apical dominance Increased tolerance to drought
13	G156	KO	Altered seed Growth regulation ; nutrient uptake	Seed color alteration Altered C/N sensing
15	G157	OE	Altered flowering time	Altered flowering time (modest level of overexpression triggers early flowering, whereas a larger increase delays flowering)
17	G162	OE	Altered seed oil Altered seed protein	Increased seed oil content Increased seed protein content
19	G175	OE	Increased tolerance to abiotic stress	Increased tolerance to osmotic stress Increased tolerance to drought
21	G180	OE	Altered seed oil Altered flowering time	Decreased seed oil Early flowering
23	G183	OE	Altered flowering time Altered light response and/or shade tolerance Growth regulation ; nutrient uptake	Early flowering Constitutive photomorphogenesis Altered C/N sensing
25	G188	KO	Increased susceptibility to disease Increased tolerance to abiotic stress	Increased susceptibility to *Fusarium* Better germination under osmotic stress Increased tolerance to drought
27	G189	OE	Altered size Growth regulation ; nutrient uptake	Increased leaf size Altered C/N sensing
29	G192	OE	Altered flowering time Altered seed oil	Late flowering Decreased seed oil content
31	G196	OE	Increased tolerance to abiotic stress	Increased tolerance to high salt
33	G211	OE	Altered leaf biochemistry Altered architecture Altered leaf	Increase in leaf xylose Reduced apical dominance Altered leaf shape
35	G214	OE	Altered flowering time Altered leaf biochemistry Altered seed prenyl lipids Altered leaf prenyl lipids	Late flowering Increased leaf fatty acids Increased seed lutein Increased leaf chlorophyll and carotenoids
37	G226	OE	Altered seed protein Altered trichomes Altered root Increased tolerance to abiotic stress Growth regulation ; nutrient uptake	Increased seed protein Glabrous, lack of trichomes Increased root hairs Increased tolerance to high salt Increased tolerance to nitrogen-limited medium Altered C/N sensing
2267	G234	OE	Growth regulation ; nutrient uptake	Altered C/N sensing
2269	G237	OE	Growth regulation ; nutrient uptake	Altered C/N sensing
39	G241	KO	Altered seed protein Altered seed oil	Increased seed protein content Decreased seed oil

			Altered sugar sensing	Decreased germination and growth on glucose medium
41	G248	OE	Increased susceptibility to disease	Increased susceptibility to *Botrytis*
43	G254	OE	Altered sugar sensing	Decreased germination and growth on glucose medium
45	G256	OE	Increased tolerance to abiotic stress	Better germination and growth in cold
47	G278	OE	Increased susceptibility to disease	Increased susceptibility to *Sclerotinia*
49	G291	OE	Altered seed oil	Increased seed oil content
51	G303	OE	Increased tolerance to abiotic and osmotic stress	Better germination on high sucrose and high NaCl Increased tolerance to drought
53	G312	OE	Increased tolerance to abiotic stress	Better germination on high NaCl
55	G325	OE	Increased tolerance to abiotic and osmotic stress	Better germination on high sucrose and NaCl Increased tolerance to drought
57	G343	OE	Altered glyphosate sensitivity Altered size	Increased resistance to glyphosate Small plant
G2295	G347	OE	Growth regulation ; nutrient uptake	Altered C/N sensing
59	G353	OE	Increased tolerance to abiotic and osmotic stress Altered size Altered leaf Altered flower	Increased seedling vigor on polyethylene glycol (PEG) Increased tolerance to drought Reduced size Altered leaf development Short pedicels, downward pointing siliques
61	G354	OE	Altered size Altered light response and/or shade tolerance Flower	Reduced size Constitutive photomorphogenesis Short pedicels, downward pointing siliques
63	G361	OE	Altered flowering time	Late flowering
65	G362	OE	Altered flowering time Altered size Altered trichomes Morphology: other	Late flowering Reduced size Ectopic trichome formation, increased trichome number Increased pigmentation in seed and embryos, and in other organs
67	G371	OE	Increased susceptibility to disease	Increased susceptibility to *Botrytis*
69	G390	OE	Altered architecture	Altered shoot development
71	G391	OE	Altered architecture	Altered shoot development
73	G409	OE	Increased tolerance to disease	Increased tolerance to *Erysiphe*
75	G427	OE	Altered seed oil Altered seed protein Growth regulation; nutrient uptake	Increased oil content Decreased protein content Altered C/N sensing
77	G438	KO	Altered stem morphology Altered architecture	Reduced lignin Reduced branching
79	G450	OE	Altered seed	Increased seed size

81	G464	OE	Increased tolerance to abiotic stress	Better germination and growth in heat
83	G470	OE	Altered fertility	Short stamen filaments
85	G477	OE	Increased susceptibility to disease Increased tolerance to abiotic stress	Increased susceptibility to *Sclerotinia* Increased sensitivity to oxidative stress
87	G481	OE	Increased tolerance to abiotic and osmotic stress	Altered sugar sensing: better germination on sucrose media Increased tolerance to drought
89	G482	OE	Increased tolerance to abiotic and osmotic stress	Increased tolerance to high salt
91	G484	KO	Altered seed glucosinolates	Altered glucosinolate profile
93	G489	OE	Increased tolerance to abiotic and osmotic stress	Increased tolerance to osmotic stress Increased tolerance to drought
95	G490	OE	Altered flowering time	Early flowering
97	G504	OE	Altered seed oil composition	Decreased seed oil composition and content; increase in 18:2 fatty acid and decrease in 20:1 fatty acid
99	G509	KO	Altered seed oil Altered seed protein	Increased total seed oil and protein content
101	G519	OE	Altered seed oil	Increased seed oil content
103	G545	OE	Increased tolerance to abiotic and osmotic stress Increased susceptibility to disease Growth regulation; nutrient uptake	Susceptible to high salt Increased susceptibility to *Erysiphe* Increased susceptibility to *Pseudomonas* Increased susceptibility to *Fusarium* Increased tolerance to phosphate-free medium Altered C/N sensing
105	G546	OE	Altered hormone sensitivity	Decreased sensitivity to abscisic acid (ABA)
107	G561	OE	Altered seed oil Tolerance to abiotic stress	Increased seed oil content Increased tolerance to potassium-free medium
109	G562	OE	Altered flowering time	Late flowering
111	G567	OE	Altered seed oil Altered seed protein Altered sugar sensing	Increased total seed oil/ protein content Increased total seed oil/ protein content Decreased seedling vigor on high glucose
113	G568	OE	Altered architecture	Altered branching
115	G584	OE	Altered seed	Large seeds
117	G585	OE	Altered trichomes	Reduced trichome density
119	G590	KO OE OE	Altered seed oil Altered flowering time Growth regulation; nutrient uptake	Increased seed oil content Early flowering Altered C/N sensing
121	G594	OE	Increased susceptibility to disease	Increased susceptibility to *Sclerotinia*
123	G597	OE	Altered seed protein	Altered seed protein content
125	G598	OE	Altered seed oil	Increased seed oil
127	G634	OE	Altered trichomes Altered light response and/or shade tolerance	Increased trichome density and size Increased shade tolerance; lack of shade avoidance phenotype

			Increased tolerance to abiotic stress	Increased drought tolerance
129	G635	OE	Variegation Growth regulation; nutrient uptake	Altered coloration Altered C/N sensing
131	G636	OE	Altered senescence	Premature senescence
133	G638	OE	Altered flower	Altered flower development
135	G652	KO	Altered seed prenyl lipids	Increase in alpha-tocopherol
2391	G657	OE	Growth regulation; nutrient uptake	Altered C/N sensing
137	G663	OE	Altered pigment	Increased anthocyanins in leaf, root, seed
139	G664	OE	Increased tolerance to abiotic stress	Better germination and growth in cold
141	G674	OE	Altered leaf	Dark green, upwardly oriented leaves
143	G676	OE	Altered trichomes	Reduced trichome number, ectopic trichome formation
145	G680	OE	Altered sugar sensing	Reduced germination on glucose medium
147	G682	OE	Altered trichomes Increased tolerance to abiotic and osmotic stress Altered root Growth regulation; nutrient uptake	Glabrous, lack of trichomes Better germination and growth in heat Increased root hairs Increased tolerance to drought Altered C/N sensing
149	G715	OE	Altered seed oil	Increased seed oil content
151	G720	OE KO	Increased tolerance to abiotic and osmotic stress Increased susceptibility to abiotic and osmotic stress	More freezing tolerant Increased susceptibility to freezing
153	G736	OE	Altered flowering time Altered leaf	Late flowering Altered leaf shape
155	G748	OE	Altered seed prenyl lipids Altered stem morphology Altered flowering time	Increased lutein content More vascular bundles in stem Late flowering
2413	G760	OE	Growth regulation; nutrient uptake	Altered C/N sensing
157	G779	OE	Altered fertility Altered flower	Reduced fertility Homeotic transformations
159	G789	OE	Altered flowering time	Early flowering
161	G801	OE	Increased tolerance to abiotic stress	Better germination on high NaCl
163	G849	KO	Altered seed oil Altered seed protein	Increased seed oil content Altered seed protein content
165	G859	OE	Altered flowering time	Late flowering
167	G864	OE	Increased tolerance to abiotic stress	Better germination in heat Increased tolerance to drought
169	G867	OE	Increased tolerance to abiotic and osmotic stress Altered sugar sensing	Better seedling vigor on high salt Better seedling vigor on high sucrose
171	G869	OE	Altered seed oil composition	Altered seed fatty acid composition
2437	G872	OE	Growth regulation; nutrient uptake	Altered C/N sensing
173	G877	KO	Embryo lethal	Embryo lethal phenotype: potential

				herbicide target
175	G881	OE	Increased susceptibility to disease	Increased susceptibility to *Erysiphe*
177	G892	KO	Altered seed protein Altered seed oil	Altered seed protein content Altered seed oil content
179	G896	KO	Increased susceptibility to disease	Increased susceptibility to *Fusarium*
2451	G904	OE	Growth regulation; nutrient uptake	Altered C/N sensing
181	G910	OE	Altered flowering time	Late flowering
183	G911	OE	Tolerance to abiotic stress	Increased growth on potassium-free medium
185	G912	OE	Increased tolerance to abiotic stress Altered pigment Altered sugar sensing	Freezing tolerant Increased survival in drought conditions Dark green color Reduced cotyledon expansion in glucose
187	G913	OE	Increased tolerance to abiotic stress Altered flowering time	Increased tolerance to freezing Late flowering Increased tolerance to drought
189	G922	OE	Increased tolerance to abiotic and osmotic stress	Better germination on high sucrose Better germination, increased root growth on high salt Increased tolerance to drought
191	G926	KO	Altered hormone sensitivity Increased tolerance to abiotic and osmotic stress	Reduced sensitivity to ABA Increased tolerance to high salt and sucrose
2065	G932	OE	Growth regulation; nutrient uptake	C/N sensing: increased tolerance to low nitrogen
2461	G937	OE	Growth regulation; nutrient uptake	Altered C/N sensing
2471	G958	OE	Growth regulation; nutrient uptake	Altered C/N sensing
193	G961	KO	Altered seed oil	Increased seed oil content
2479	G964	KO	Growth regulation; nutrient uptake	Altered C/N sensing
195	G971	OE	Altered flowering time	Late flowering
197	G974	OE	Altered seed oil	Altered seed oil content
199	G975	OE	Altered leaf biochemistry Increased tolerance to abiotic and osmotic stress Growth regulation; nutrient uptake	Increased fatty acids and wax in leaves Increased tolerance to drought Altered C/N sensing
201	G979	KO	Altered seed Growth regulation; nutrient uptake	Altered seed development, ripening, and germination Altered C/N sensing
203	G987	KO	Altered leaf fatty acids Altered leaf biochemistry	Reduction in 16:3 fatty acids Altered prenyl lipids: chlorophyll, tocopherol, carotenoid
205	G988	OE	Altered seed protein Altered flower Altered architecture Altered stem morphology	Increased seed protein content Enlarged floral organs, short pedicels Reduced lateral branching Thicker stem, altered distribution of

			Growth regulation; nutrient uptake	vascular bundles Altered C/N sensing
207	G1040	OE	Altered seed	Smaller and more rounded seeds
209	G1047	OE	Increased tolerance to disease	Increased tolerance to *Fusarium*
2515	G1048	OE	Altered light response and/or shade tolerance	Increased shade tolerance; lack of shade avoidance phenotype
2517	G1049	OE	Growth regulation; nutrient uptake	Altered C/N sensing
211	G1051	OE	Altered flowering time	Late flowering
213	G1052	OE	Altered flowering time	Late flowering
215	G1062	KO	Altered seed	Altered seed shape
217	G1063	OE	Altered leaf Altered inflorescence Altered flower	Altered leaf shape, dark green color Altered inflorescence development Altered flower development, ectopic carpel tissue
219	G1064	OE	Increased susceptibility to disease	Increased sensitivity to *Botrytis*
221	G1069	OE	Altered hormone sensitivity Increased tolerance to abiotic and osmotic stress Growth regulation; nutrient uptake	Reduced ABA sensitivity Better germination under osmotic stress Increased tolerance to drought Altered C/N sensing
223	G1073	OE	Altered size Altered seed Increased tolerance to abiotic stress	Substantially increased plant size Increased seed yield Increased tolerance to drought
225	G1075	OE	Altered flower	Reduced or absent petals, sepals and stamens
227	G1084	OE	Increased susceptibility to disease	Increased susceptibility to *Botrytis*
229	G1089	KO	Increased tolerance to osmotic stress	Better germination under osmotic stress
231	G1134	OE	Altered hormone sensitivity	Altered response to ethylene: longer hypocotyls and lack of apical hook
233	G1140	OE	Altered flower	Altered flower development
235	G1143	OE	Altered seed oil	Altered seed oil content
237	G1146	OE	Altered leaf	Altered leaf development
239	G1196	KO	Increased susceptibility to disease	Increased susceptibility to *Botrytis*
241	G1198	OE	Altered seed oil	Increased seed oil content
243	G1225	OE	Altered flowering time Altered sugar sensing	Early flowering Better germination on sucrose and glucose media
245	G1226	OE	Altered seed oil	Increased seed oil content
247	G1229	OE	Altered seed oil	Decreased seed oil content
2555	G1246	OE	Growth regulation; nutrient uptake	Altered C/N sensing
249	G1255	OE	Increased susceptibility to disease Altered seed Altered architecture Growth regulation; nutrient uptake	Increased susceptibility to *Botrytis* Increased seed size Reduced apical dominance Altered C/N sensing

251	G1266	OE	Increased tolerance to disease Growth regulation; nutrient uptake	Increased tolerance to *Erysiphe* Altered C/N sensing
253	G1275	OE	Altered architecture	Reduced apical dominance
255	G1305	OE	Increased tolerance to abiotic stress	Reduced chlorosis in heat
257	G1322	OE	Increased tolerance to abiotic stress Altered size Leaf glucosinolates Altered light response and/or shade tolerance Growth regulation; nutrient uptake	Increased seedling vigor in cold Reduced size Increase in M39480 Constitutive photomorphogenesis Altered C/N sensing: increased tolerance to low nitrogen
259	G1323	OE	Altered seed oil Altered seed protein	Decreased seed oil Increased seed protein
261	G1330	OE	Altered hormone sensitivity	Ethylene insensitive when germinated in the dark on ACC
263	G1331	OE	Altered light response and/or shade tolerance Growth regulation; nutrient uptake	Constitutive photomorphogenesis Altered C/N sensing
265	G1332	OE	Altered trichomes Growth regulation; nutrient uptake	Reduced trichome density Altered C/N sensing
267	G1363	OE	Increased tolerance to disease	Increased tolerance to *Fusarium*
269	G1411	OE	Altered architecture	Loss of apical dominance
2607	G1412	KO	Altered light response and/or shade tolerance	Increased shade tolerance; lack of shade avoidance phenotype
271	G1417	KO	Altered seed oil	Increase in 18:2, decrease in 18:3 fatty acids
273	G1419	OE	Altered seed protein	Increased seed protein
275	G1449	OE	Altered flower	Altered flower structure
277	G1451	OE OE KO	Altered size Altered leaf Altered seed oil	Increased plant size Large leaf size Altered seed oil content
279	G1452	OE	Altered trichomes Altered leaf Altered hormone sensitivity Altered flowering time Increased tolerance to abiotic and osmotic stress	Reduced trichome density Altered leaf shape, dark green color Reduced sensitivity to ABA Better germination on sucrose, salt Late flowering Increased tolerance to drought
281	G1463	OE	Altered senescence	Premature senescence
283	G1471	OE	Altered seed oil	Increased seed oil content
285	G1478	OE	Altered seed protein Altered flowering time Altered seed oil	Decreased seed protein content Late flowering Increased seed oil content
287	G1482	KO OE	Altered pigment Altered root	Increased anthocyanins Increased root growth
289	G1488	OE	Altered seed protein Altered light response and/or shade tolerance Altered architecture	Altered seed protein content Constitutive photomorphogenesis Reduced apical dominance, shorter stems
291	G1494	OE	Altered flowering time	Early flowering

			Altered light response and/or shade tolerance Altered leaf Growth regulation; nutrient uptake	Long hypocotyls, altered leaf shape Pale green leaves, altered leaf shape Altered C/N sensing
293	G1496	OE	Altered seed oil	Altered seed oil content
295	G1499	OE	Altered pigment Altered architecture Altered flower	Dark green color Altered plant architecture Altered floral organ identity and development
297	G1519	KO	Embryo lethal	Embryo lethal phenotype: potential herbicide target
299	G1526	KO	Altered seed oil	Increased seed oil content
301	G1540	OE	Altered cell differentiation	Reduced cell differentiation in meristem
303	G1543	OE	Altered architecture Morphology: other Altered seed oil Altered leaf prenyl lipids	Altered architecture, compact plant Dark green color Decreased seed oil Increase in chlorophyll a and b
2667	G1587	OE	Growth regulation; nutrient uptake	Altered C/N sensing
305	G1634	OE	Altered seed oil Altered seed protein	Increased seed oil content Decreased seed protein content
307	G1637	OE	Altered seed protein	Altered seed protein content
309	G1640	OE	Altered seed oil	Increased seed oil
311	G1645	OE	Altered inflorescence	Altered inflorescence structure
313	G1646	OE	Altered seed oil	Increased seed oil content
2685	G1649	OE	Growth regulation; nutrient uptake	Altered C/N sensing
315	G1652	OE	Altered seed protein	Increased seed protein content
	G1666	KO	Growth regulation; nutrient uptake	Altered C/N sensing
317	G1672	OE	Altered seed oil	Altered seed oil content
319	G1677	OE	Altered seed protein Altered seed oil	Altered seed protein content Altered seed oil content
321	G1749	OE	Altered necrosis	Formation of necrotic lesions
323	G1750	OE	Altered seed oil Growth regulation; nutrient uptake	Increased seed oil content Altered C/N sensing
325	G1756	OE	Increased susceptibility to disease	Increased susceptibility to *Botrytis*
327	G1765	OE	Altered seed oil	Increased seed oil content
329	G1777	OE	Altered seed oil Altered seed protein	Increased seed oil content Decreased seed protein content
331	G1792	OE	Altered leaf Increased tolerance to disease Increased tolerance to abiotic and osmotic stress	Dark green, shiny leaves Increased resistance to *Erysiphe* Increased resistance to *Botrytis* Increased resistance to *Fusarium* Increased tolerance to nitrogen-limited medium Increased tolerance to drought
333	G1793	OE	Altered seed oil	Increased seed oil content
335	G1794	OE	Altered architecture Altered light response and/or	Altered architecture, bushier plant Reduced apical dominance

			shade tolerance Increased tolerance to osmotic and abiotic stress	Constitutive photomorphogenesis Increased sensitivity to high PEG Reduced root growth
337	G1804	OE	Altered flowering time Altered sugar sensing	Late flowering Altered sugar sensing: more sensitive to glucose in germination assays
339	G1818	OE	Altered seed protein	Increased protein content
341	G1820	OE	Altered flowering time Altered hormone sensitivity Altered seed protein Increased tolerance to abiotic and osmotic stress	Early flowering Reduced ABA sensitivity Increased seed protein content Better germination in high NaCl Increased tolerance to drought
2733	G1835	OE	Growth regulation; nutrient uptake	Altered C/N sensing
343	G1836	OE	Increased tolerance to abiotic and osmotic stress	Better germination in high salt Increased tolerance to drought
345	G1838	OE	Altered seed oil	Increased seed oil content
347	G1841	OE	Increased tolerance to abiotic and osmotic stress Altered flowering time	Better germination under heat stress Early flowering
349	G1842	OE	Altered flowering time	Early flowering
351	G1843	OE	Altered flowering time	Early flowering
353	G1852	OE	Increased tolerance to abiotic and osmotic stress	Better root growth under osmotic stress
355	G1863	OE	Altered leaf	Altered leaf shape and coloration
357	G1880	KO	Increased tolerance to disease	Increased resistance to *Botrytis*
359	G1895	OE	Altered flowering time	Late flowering
361	G1902	OE	Altered seed oil	Increased seed oil content
363	G1903	OE	Altered seed protein	Decreased seed protein content
365	G1919	OE	Increased tolerance to disease	Increased tolerance to *Botrytis*
367	G1927	OE	Increased tolerance to disease	Increased tolerance to *Sclerotinia*
369	G1930	OE	Increased tolerance to osmotic stress Growth regulation; nutrient uptake	Better germination under osmotic stress Altered C/N sensing
371	G1936	KO	Increased susceptibility to disease	Increased susceptibility to *Sclerotinia* Increased susceptibility to *Botrytis*
373	G1944	OE	Altered senescence	Early senescence
375	G1946	OE	Altered seed oil Altered seed protein Altered flowering time Growth regulation; nutrient uptake	Increased seed oil content Decreased seed protein content Early flowering Increased root growth on phosphate-free media
377	G1947	KO	Altered fertility	Reduced fertility
379	G1948	OE	Altered seed oil	Increased seed oil content
381	G1950	OE	Increased tolerance to disease	Increased tolerance to *Botrytis*
383	G1958	KO	Altered size Altered seed oil Altered seed protein	Reduced size and root mass Increased seed oil content Increased seed protein content.
2157	G1995	OE	Altered light response and/or shade tolerance	Increased shade tolerance; lack of shade avoidance phenotype
385	G2007	OE	Altered flowering time	Late flowering

387	G2010	OE	Altered flowering time	Early flowering
389	G2053	OE	Increased tolerance to abiotic and osmotic stress Growth regulation; nutrient uptake	Increased root growth under osmotic stress Increased tolerance to drought Altered C/N sensing
2797	G2057	OE	Growth regulation; nutrient uptake	Altered C/N sensing
391	G2059	OE	Altered seed oil Altered seed protein	Altered seed oil content Altered seed protein content
393	G2085	OE	Altered seed	Increased seed size and altered seed color
395	G2105	OE	Altered seed	Large, pale seeds
397	G2110	OE	Increased tolerance to abiotic and osmotic stress	Increased tolerance to high salt Increased tolerance to drought
399	G2114	OE	Altered seed	Increased seed size
401	G2117	OE	Altered seed protein	Increased seed protein content Altered C/N sensing
403	G2123	OE	Altered seed oil	Increased seed oil content
405	G2130	OE	Increased tolerance to abiotic stress	Better germination in heat
2163	G2131	OE	Growth regulation; nutrient uptake	Altered C/N sensing
407	G2133	OE	Altered herbicide sensitivity Altered flowering time Increased tolerance to abiotic and osmotic stress Growth regulation; nutrient uptake	Increased tolerance to glyphosate Late flowering Increased drought tolerance Altered C/N sensing
409	G2138	OE	Altered seed oil	Increased seed oil content
411	G2140	OE	Altered hormone sensitivity Increased tolerance to abiotic and osmotic stress	Decreased sensitivity to ABA Better germination on high NaCl and sucrose Increased tolerance to drought
413	G2143	OE	Altered inflorescence Altered leaf Altered flower	Altered inflorescence development Altered leaf shape, dark green color Altered flower development, ectopic carpel tissue
415	G2144	OE	Altered flowering time Altered leaf Light response Growth regulation; nutrient uptake	Early flowering Pale green leaves, altered leaf shape Long hypocotyls, altered leaf shape Altered C/N sensing
2827	G2145	OE	Growth regulation; nutrient uptake	Altered C/N sensing
417	G2153	OE	Increased tolerance to osmotic stress	Better germination under osmotic stress
419	G2155	OE	Altered flowering time	Late flowering
421	G2192	OE	Altered seed oil	Altered seed fatty acid composition
423	G2295	OE	Altered flowering time	Early flowering
425	G2340	OE	Altered seed glucosinolates	Altered glucosinolate profile
427	G2343	OE	Altered seed oil	Increased seed oil content
429	G2346	OE	Altered size	Enlarged seedlings
431	G2347	OE	Altered flowering time	Early flowering
433	G2379	OE	Increased tolerance to osmotic	Increased seedling vigor on high

			stress	sucrose media
435	G2430	OE	Increased tolerance to abiotic and osmotic stress Altered size	Increased tolerance to heat Increased leaf size, faster development
437	G2505	OE	Increased tolerance to abiotic and osmotic stress Altered light response and/or shade tolerance	Increased tolerance to drought Increased shade tolerance; lack of shade avoidance phenotype
439	G2509	OE	Altered seed oil Altered seed protein Altered seed prenyl lipids Altered architecture Altered flowering time	Decreased seed oil content Increased seed protein content Increase in alpha-tocopherol Reduced apical dominance Early flowering
2875	G2512	OE	Growth regulation; nutrient uptake	Altered C/N sensing
441	G2517	OE	Altered herbicide sensitivity	Increased tolerance to glyphosate
443	G2520	OE	Altered seed prenyl lipids Growth regulation; nutrient uptake	Altered tocopherol composition Altered C/N sensing
2185	G2535	OE	Growth regulation; nutrient uptake	Altered C/N sensing
445	G2555	OE	Altered light response and/or shade tolerance Increased susceptibility to disease	Constitutive photomorphogenesis Increased susceptibility to *Botrytis*
447	G2557	OE	Altered leaf Altered flower	Altered leaf shape, dark green color Altered flower development, ectopic carpel tissue
449	G2583	OE	Altered leaf	Glossy, shiny leaves
451	G2701	OE	Increased tolerance to abiotic and osmotic stress	Better germination on high NaCl and sucrose Increased tolerance to drought
2191	G2718	OE	Growth regulation; nutrient uptake	Altered C/N sensing
453	G2719	OE	Increased tolerance to osmotic stress Growth regulation; nutrient uptake	Increased seedling vigor on high sucrose Altered C/N sensing
2193	G2776	OE	Increased tolerance to abiotic and osmotic stress	Increased tolerance to drought
455	G2789	OE	Altered hormone sensitivity Increased tolerance to abiotic and osmotic stress Altered light response and/or shade tolerance Growth regulation; nutrient uptake	Better germination on high sucrose Reduced ABA sensitivity Increased tolerance to drought Increased shade tolerance; lack of shade avoidance phenotype Altered C/N sensing
457	G2830	KO	Altered seed oil	Increased seed oil content
1951	G12	KO OE	Altered hormone sensitivity Altered necrosis	Increased sensitivity to ACC Leaf and hypocotyl necrosis
1953	G30	OE	Altered leaf Altered light response and/or shade tolerance	Glossy green leaves Increased shade tolerance; lack of shade avoidance phenotype
1975	G231	OE	Altered leaf biochemistry	Increased leaf unsaturated fatty acids

			Altered seed oil Altered seed protein	Increased seed oil content Decreased seed protein content
1979	G247	OE	Altered trichomes	Altered trichome distribution, reduced trichome density
1991	G370	KO OE	Altered size Altered trichome	Reduced size, shiny leaves ectopic trichome formation
2009	G485	OE KO	Altered flowering time Altered flowering time	Early flowering Late flowering
2061	G839	OE	Growth regulation; nutrient uptake	Increased tolerance to nitrogen-limited medium
2099	G1357	OE	Altered leaf Increased tolerance to abiotic and osmotic stress Altered hormone sensitivity Altered flowering time	Altered leaf shape, dark green leaves Increased tolerance to cold Insensitive to ABA Late flowering
2126	G1646	OE	Altered seed oil	Increased seed oil content
2142	G1816	OE	Altered sugar sensing Growth regulation; nutrient uptake Increased tolerance to abiotic and osmotic stress Altered root Altered trichomes	Increased tolerance to glucose Altered C/N sensing; less anthocyanin on nitrogen-limited medium Increased tolerance to osmotic stress Increased root hairs Glabrous leaves Increased tolerance to nitrogen-limited medium
2147	G1888	OE	Altered size	Reduced size, dark green leaves
2153	G1945	OE	Altered flowering time Altered leaf	Late flowering Altered leaf shape
2195	G2826	OE	Altered flower Altered trichomes	Aerial rosettes Ectopic trichome formation
2197	G2838	OE	Altered trichomes Altered flowering time Altered flower Leaves Altered size	Increased trichome density Late flowering Flower: multiple alterations Aerial rosettes Dark green leaves Increased seedling size
2199	G2839	OE	Increased tolerance to abiotic and osmotic stress Altered inflorescence Altered size	Better germination on high sucrose Increased tolerance to drought Downward pedicels Reduced size

Table 8 shows the polypeptides identified by SEQ ID NO; Gene ID (GID) No.; the transcription factor family to which the polypeptide belongs, and conserved domains of the polypeptide. The first column shows the polypeptide SEQ ID NO; the third column shows the transcription factor family to which the polynucleotide belongs; and the fourth column shows the amino acid residue positions of the conserved domain in amino acid (AA) co-ordinates.

Table 8. Gene families and conserved domains

Polypeptide SEQ ID NO:	GID No.	Family	Conserved Domains in Amino Acid Coordinates
4	G19	AP2	76-145
6	G22	AP2	89-157
10	G28	AP2	145-213
12	G47	AP2	11-80
38	G226	MYB-related	34-69
60	G353	Z-C2H2	41-61, 84-104
62	G354	Z-C2H2	42-62, 88-109
88	G481	CAAT	20-110
90	G482	CAAT	26-116
2010	G485	CAAT	20-110
94	G489	CAAT	57-156
128	G634	TH	62-147, 189-245
148	G682	MYB-related	27-63
168	G864	AP2	119-186
170	G867	AP2	59-124
186	G912	AP2	51-118
188	G913	AP2	62-128
190	G922	SCR	225-242
192	G926	CAAT	131-225
200	G975	AP2	4-71
2516	G1048	bZIP	138-190
222	G1069	AT-hook	67-74
224	G1073	AT-hook	33-42, 78-175
226	G1075	AT-hook	78-85
2102	G1364	CAAT	29-119
270	G1411	AP2	87-154
278	G1451	ARF	22-357
304	G1543	HB	135-195
332	G1792	AP2	17-85
2142	G1816	MYB-related	26-61
342	G1820	CAAT	70-133
344	G1836	CAAT	30-164
370	G1930	AP2	59-124
2608	G1995	Z-C2H2	93-113
408	G2133	AP2	11-83
418	G2153	AT-hook	75-94, 162-206
420	G2155	AT-hook	18-38
2172	G2345	CAAT	28-118
440	G2509	AP2	89-156
450	G2583	AP2	4-71
	G2718	MYB-related	28-64

Examples of some of the utilities that may be desirable in plants, and that may be provided by transforming the plants with the presently disclosed sequences, are listed in Table 9. Many of the transcription factors listed in Table 9 may be operably linked with a specific promoter that causes the

transcription factor to be expressed in response to environmental, tissue-specific or temporal signals. For example, G362 induces ectopic trichomes on flowers but also produces small plants. The former may be desirable to produce insect or herbivore resistance, or increased cotton yield, but the latter may be undesirable in that it may reduce biomass. However, by operably linking G362 with a flower-specific promoter, one may achieve the desirable benefits of the gene without affecting overall biomass to a significant degree. For examples of flower specific promoters, see Kaiser et al. (*supra*). For examples of other tissue-specific, temporal-specific or inducible promoters, see the above discussion under the heading "Vectors, Promoters, and Expression Systems".

Table 9. Genes, traits and utilities that affect plant characteristics

Trait Category	Phenotype(s)	Transcription factor genes that impact traits	Utility
Abiotic stress	Effect of chilling on plants Increased tolerance:	G256; G664; G1322	Improved germination, growth rate, earlier planting, yield
	Germination in cold Increased tolerance:	G256; G664	Earlier planting; improved survival, yield
	Freezing tolerance	G720 (G720 KO is more susceptible); G912; G913	Earlier planting; improved quality, survival, yield
	Drought Increased tolerance:	G47; G175; G188; G303; G325; G353; G481; G489; G634; G682; G864; G912; G913; G922; G975; G1069; G1452; G1792; G1820; G1836; G2053; G2110; G2133; G2140; G2505; G2701; G2776; G2789; G2839	Improved survival, vigor, appearance, yield
	Heat Increased tolerance:	G464; G682; G864; G1305; G1841; G2130; G2430	Improved germination, growth rate, later planting, yield
	Osmotic stress Increased sensitivity:	G1794	Abiotic stress response manipulation
	Increased tolerance:	G47; G175; G188; G303; G325; G353; G489; G922; G926; G1069; G1089; G1452; G1816; G1820; G1852; G1930; G2053;	Improved germination rate, seedling vigor, survival, yield

		G2140; G2153; G2379; G2701; G2719; G2789; G2839	
	Salt tolerance More susceptible:	G545	Manipulation of response to high salt conditions
	Increased tolerance:	G22; G196; G226; G312; G482; G801; G867; G922; G1836; G2110	Improved germination rate, survival, yield; extended growth range
	Nitrogen stress Sensitivity to N limitation:	G1794	Manipulation of response to low nutrient conditions
	Tolerance to N limitation:	G225; G226; G839; G1792; G1816	Improved yield and nutrient stress tolerance, decreased fertilizer usage
	Phosphate stress Tolerance to P limitation:	G545; G561; G911; G1946	Improved yield and nutrient stress tolerance, decreased fertilizer usage
	Oxidative stress	G477	Improved yield, quality, ultraviolet and chemical stress tolerance
Herbicide	Glyphosate	G343; G2133; G2517	Generation of glyphosate-resistant plants to improve weed control
Hormone sensitivity	Abscisic acid (ABA) sensitivity Reduced sensitivity to ABA:	G546; G926; G1069; G1357; G1452; G1820; G2140; G2789	Modification of seed development, improved seed dormancy, cold and dehydration tolerance
	Sensitivity to ethylene Altered response:	G1134	Manipulation of fruit ripening
	Insensitive to ethylene:	G1330	
Disease	*Botrytis* Increased susceptibility:	G248; G371; G1064; G1084; G1196; G1255; G1756; G1936; G2555	Manipulation of response to disease organism
	Increased resistance or tolerance:	G28; G1792; G1880; G1919; G1950	Improved yield, appearance, survival,

			extended range
	Fusarium		
	Increased susceptibility:	G188; G545; G896	Manipulation of response to disease organism
	Increased resistance or tolerance:	G1047; G1792	Improved yield, appearance, survival, extended range
	Erysiphe		
	Increased susceptibility:	G545; G881	Manipulation of response to disease organism
	Increased resistance or tolerance:	G19; G28; G409; G1266; G1363; G1792	Improved yield, appearance, survival, extended range
	Pseudomonas		
	Increased susceptibility:	G545	Manipulation of response to disease organism
	Sclerotinia		
	Increased susceptibility:	G278; G477; G594; G1936	Manipulation of response to disease organism
	Increased resistance or tolerance:	G28; G1927	Improved yield, appearance, survival, extended range
Growth regulator	Altered sugar sensing		Alteration of energy balance, photosynthetic rate, carbohydrate accumulation, biomass production, source-sink relationships, senescence; alteration of storage compound accumulation in seeds
	Decreased tolerance to sugars:	G241; G254; G567; G680; G912; G1804	
		G481; G867; G1225; G1816	
	Increased tolerance to sugars:		
	Altered C/N sensing	G682; G226; G1816; G2718; G24; G545; G760; G937; G971; G988; G1069; G1322; G1587; G1666; G2117; G2131; G2520; G2789; G8; G27; G156; G183; G189; G234; G237; G347; G427; G590; G635; G657; G872; G904; G912; G932; G958; G964; G975; G979; G1049; G1246; G1255; G1266; G1331; G1332; G1494; G1649; G1750; G1816;G1835; G1930;	

		G2053; G2057; G2133; G2144; G2145; G2295; G2512; G2535; G2719	
Flowering time	Early flowering	G157; G180; G183; G485 (OE); G490; G590; G789; G1225; G1494; G1820; G1841; G1842; G1843; G1946; G2010; G2144; G2295; G2347; G2509	Faster generation time; synchrony of flowering; additional harvests within a growing season, shortening of breeding programs
	Late flowering	G8; G47; G157; G192; G214; G231; G361; G362; G485 (KO); G562; G736; G748; G859; G910; G913; G971; G1051; G1052; G1357; G1452; G1478; G1804; G1895; G1945; G2007; G2133; G2155; G2838	Increased yield or biomass, alleviate risk of transgenic pollen escape, synchrony of flowering
General development and morphology	Altered flower structure Stamen: Sepal: Petal: Pedicel: Carpel: Multiple alterations: Enlarged floral organs: Siliques: Reduced fertility: Aerial rosettes	G988; G1075; G1140; G1499; G2557 G1075; G1140; G2557 G638; G1075; G1140; G1449; G1499; G2557 G353; G354; G988 G1063; G1140; G2143; G2143; G2557 G638; G988; G1063; G1140; G1449; G1499; G2143; G2557 G988; G1449; G2838 G353; G354 G470; G779; G988; G1075; G1140; G1499; G1947; G2143; G2557 G638; G779; G1140; G1499 G1995; G2826; G2838	Ornamental modification of plant architecture, improved or reduced fertility to mitigate escape of transgenic pollen, improved fruit size, shape, number or yield
	Inflorescence architectural change Altered branching pattern: Short internodes/bushy inflorescences: Internode elongation: Lack of inflorescence:	G47; G1063; G1645; G2143 G47 G1063 G1499; G2143	Ornamental modification of flower architecture; timing of flowering; altered plant habit for yield or harvestability benefit, reduction in pollen production of genetically modified plants; manipulation of seasonality and annual or perennial habit; manipulation of determinate vs.

			indeterminate growth
	Altered shoot meristem development Stem bifurcations:	G390; G391	Ornamental modification of plant architecture, manipulation of growth and development, increase in leaf numbers, modulation of branching patterns to provide improved yield or biomass
	Altered branching pattern	G427; G568; G988; G1543; G1794	Ornamental modification of plant architecture, improved lodging resistance
	Apical dominance Reduced apical dominance:	G47; G211; G1255; G1275; G1411; G1488; G1794; G2509	Ornamental modification of plant architecture
	Altered trichome density; development, or structure Reduced or no trichomes: Ectopic trichomes/altered trichome development/cell fate: Increase in trichome number, size or density:	G225; G226; G247; G585; G676; G682; G1332; G1452; G1816 G247; G362; G370; G676; G2826 G362; G634; G838; G2838	Ornamental modification of plant architecture, increased plant product (e.g., diterpenes, cotton) productivity, insect and herbivore resistance
	Stem morphology and altered vascular tissue structure	G47; G438; G748; G988; G1488	Modulation of lignin content; improvement of wood, palatability of fruits and vegetables
	Root development Increased root growth and proliferation: Increased root hairs:	G1482 G225; G226; G1816	Improved yield, stress tolerance; anchorage
	Altered seed development, ripening and germination	G979	
	Cell differentiation and cell proliferation	G1540	Increase in carpel or fruit development; improve regeneration

			of shoots from callus in transformation or micro-propagation systems
	Rapid development	G2430	Promote faster development and reproduction in plants
	Senescence Premature senescence:	G636; G1463; G1944	Improvement in response to disease, fruit ripening
	Lethality when overexpressed	G877; G1519	Herbicide target; ablation of specific tissues or organs such as stamen to prevent pollen escape
	Necrosis	G12, G24	Disease resistance
Plant size	Increased plant size	G1073; G1451	Improved yield, biomass, appearance
	Larger seedlings	G2346; G2838	Increased survival and vigor of seedlings, yield
	Dwarfed or more compact plants	G24; G343; G353; G354; G362; G370; G1008; G1277; G1543; G1794; G1958	Dwarfism, lodging resistance, manipulation of gibberellin responses
Leaf morphology	Dark green leaves	G674; G912; G1063; G1357; G1452; G1482; G1499; G1792; G1863; G1888; G2143; G2557; G2838	Increased photosynthesis, biomass, appearance, yield
	Change in leaf shape	G211; G353; G674; G736; G1063; G1146; G1357; G1452; G1494; G1543; G1863; G2143; G2144	Ornamental applications
	Altered leaf size: Increased leaf size, number or mass:	G189; G214; G1451; G2430	Increased yield, ornamental applications
	Light green leaves	G1494; G2144	Ornamental applications
	Variegation	G635	Ornamental applications
	Glossy leaves	G30; G1792; G2583	Ornamental applications, manipulation of wax composition, amount, or distribution
Seed morphology	Altered seed coloration	G156; G2105; G2085	Appearance

	Seed size and shape Increased seed size: Decreased seed size: Altered seed shape:	G450; G584; G1255; G2085; G2105; G2114 G1040 G1040; G1062	Yield, appearance Appearance Appearance
Leaf biochemistry	Increased leaf wax	G975; G1792; G2583	Insect, pathogen resistance
	Leaf prenyl lipids Reduced chlorophyll: Increase in tocopherols Increased lutein content Increase in chlorophyll or carotenoids:	G987 G652; G987; G2509 G748 G214; G1543	
	Leaf insoluble sugars Increase in leaf xylose	G211	
	Increased leaf anthocyanins	G663; G1482; G1888	
	Leaf fatty acids Reduction in leaf fatty acids: Increase in leaf fatty acids:	G987 G214	
Seed biochemistry	Seed oil content Increased oil content: Decreased oil content: Altered oil content: Altered fatty acid content:	G162; G291; G427; G509; G519; G561; G590; G598; G629; G715; G849; G961; G1198; G1226; G1471; G1478; G1526; G1640; G1646; G1750; G1765; G1777; G1793; G1838; G1902; G1946; G1948; G1958; G2123; G2138; G2343; G2830 G180; G192; G241; G504; G1143; G1229; G1323; G1543; G2509 G567; G892; G974; G1451; G1496; G1646; G1672; G1677 G869; G1417; G2192	Improved oil yield Reduced caloric content
	Seed protein content Increased protein content:	G162; G226; G241; G509; G988; G1323; G1419; G1652; G1818; G1820; G1958; G2117; G2509	Improved protein yield, nutritional value Reduced caloric content

	Decreased protein content:	G427; G1478; G1777; G1903; G1946	
	Altered protein content:	G162; G567; G597; G849; G892; G1634; G1637; G1677	
	Altered seed prenyl lipid content or composition	G652; G2509; G2520	Improved antioxidant and vitamin E content
	Seed glucosinolate Altered profile:	G484; G2340	
	Increased seed anthocyanins	G362; G663	
Root Biochemistry	Increased root anthocyanins	G663	
Light response/shade tolerance	Altered cotyledon, hypocotyl, petiole development; altered leaf orientation; constitutive photomorphogenesis; photomorphogenesis in low light	G183; G354; G634; G1048; G1322; G1331; G1412; G1488; G1494; G1794; G1995, G2144; G2505; G2555; G2789;	Improved shade tolerance: potential for increased planting densities and yield enhancement
Pigment	Increased anthocyanin level	G362; G663; G1482	Enhanced health benefits, improved ornamental appearance, increased stress resistance, attraction of pollinating and seed-dispersing animals

Abbreviations: N=nitrogen P=phosphate ABA=abscisic acid C/N=carbon/nitrogen balance

Detailed description of genes, traits and utilities that affect plant characteristics

The following descriptions of traits and utilities associated with the present transcription factors offer a more comprehensive description than that provided in Table 9.

Abiotic stress, general considerations

Plant transcription factors can modulate gene expression, and, in turn, be modulated by the environmental experience of a plant. Significant alterations in a plant's environment invariably result in a change in the plant's transcription factor gene expression pattern. Altered transcription factor expression patterns generally result in phenotypic changes in the plant. Transcription factor gene product(s) in transgenic plants then differ(s) in amounts or proportions from that found in wild-type or non-transformed plants, and those transcription factors likely represent polypeptides that are used to alter the response to the environmental change. By way of example, it is well accepted in the art that analytical methods based on altered expression patterns may be used to screen for phenotypic changes in a plant far more effectively than can be achieved using traditional methods.

Abiotic stress: adult stage chilling. Enhanced chilling tolerance may extend the effective growth range of chilling sensitive crop species by allowing earlier planting or later harvest. Improved chilling tolerance may be conferred by increased expression of glycerol-3-phosphate acetyltransferase in chloroplasts (see, for example, Wolter et al. (1992) et al. *EMBO J.* 4685-4692, and Murata et al. (1992) *Nature* 356: 710-713).

Chilling tolerance could also serve as a model for understanding how plants adapt to water deficit. Both chilling and water stress share similar signal transduction pathways and tolerance/adaptation mechanisms. For example, acclimation to chilling temperatures can be induced by water stress or treatment with abscisic acid. Genes induced by low temperature include dehydrins (or LEA proteins). Dehydrins are also induced by salinity, abscisic acid, water stress, and during the late stages of embryogenesis.

Another large impact of chilling occurs during post-harvest storage. For example, some fruits and vegetables do not store well at low temperatures (for example, bananas, avocados, melons, and tomatoes). The normal ripening process of the tomato is impaired if it is exposed to cool temperatures. Transcription factor genes conferring resistance to chilling temperatures, including G256, G664, and G1322 may thus enhance tolerance during post-harvest storage.

Abiotic stress: cold germination. Several of the presently disclosed transcription factor genes confer better germination and growth in cold conditions. For example, the improved germination in cold conditions seen with G256 and G664 indicates a role in regulation of cold responses by these genes and their equivalogs. These genes might be engineered to manipulate the response to low temperature stress. Genes that would allow germination and seedling vigor in the cold would have highly significant utility in allowing seeds to be planted earlier in the season with a high rate of survival. Transcription factor genes that confer better survival in cooler climates allow a grower to move up planting time in the spring and extend the growing season further into autumn for higher crop yields. Germination of seeds and survival at temperatures significantly below that of the mean temperature required for germination of seeds and survival of non-transformed plants would increase the potential range of a crop plant into regions in which it would otherwise fail to thrive.

Abiotic stress: freezing tolerance and osmotic stress. Presently disclosed transcription factor genes, including G47, G175, G188, G303, G325, G353, G489, G922, G926, G1069, G1089, G1452, G1820, G1852, G1930, G2053, G2140, G2153, G2379, G2701, G2719, G2789, G2839 and their equivalogs, that increase germination rate and/or growth under adverse osmotic conditions, could impact survival and yield of seeds and plants. Osmotic stresses may be regulated by specific molecular control mechanisms that include genes controlling water and ion movements, functional and structural stress-induced proteins, signal perception and transduction, and free radical scavenging, and many others

(Wang et al. (2001) *Acta Hort.* (ISHS) 560: 285-292). Instigators of osmotic stress include freezing, drought and high salinity, each of which are discussed in more detail below.

In many ways, freezing, high salt and drought have similar effects on plants, not the least of which is the induction of common polypeptides that respond to these different stresses. For example, freezing is similar to water deficit in that freezing reduces the amount of water available to a plant. Exposure to freezing temperatures may lead to cellular dehydration as water leaves cells and forms ice crystals in intercellular spaces (Buchanan, *supra*). As with high salt concentration and freezing, the problems for plants caused by low water availability include mechanical stresses caused by the withdrawal of cellular water. Thus, the incorporation of transcription factors that modify a plant's response to osmotic stress or improve tolerance to (e.g., by G720, G912, G913 or their equivalogs) into, for example, a crop or ornamental plant, may be useful in reducing damage or loss. Specific effects caused by freezing, high salt and drought are addressed below.

Abiotic stress: drought and low humidity tolerance. Exposure to dehydration invokes similar survival strategies in plants as does freezing stress (see, for example, Yelenosky (1989) *Plant Physiol* 89: 444-451) and drought stress induces freezing tolerance (see, for example, Siminovitch et al. (1982) *Plant Physiol* 69: 250-255; and Guy et al. (1992) *Planta* 188: 265-270). In addition to the induction of cold-acclimation proteins, strategies that allow plants to survive in low water conditions may include, for example, reduced surface area, or surface oil or wax production. A number of presently disclosed transcription factor genes, e.g., G912, G913, G1820, G1836 and G2505 increase a plant's tolerance to low water conditions and, along with their functional equivalogs, would provide the benefits of improved survival, increased yield and an extended geographic and temporal planting range.

Abiotic stress: heat stress tolerance. The germination of many crops is also sensitive to high temperatures. Presently disclosed transcription factor genes that provide increased heat tolerance, including G464, G682, G864, G1305, G1841, G2130, G2430 and their equivalogs, would be generally useful in producing plants that germinate and grow in hot conditions, may find particular use for crops that are planted late in the season, or extend the range of a plant by allowing growth in relatively hot climates.

Abiotic stress: salt. The genes in Table 9 that provide tolerance to salt may be used to engineer salt tolerant crops and trees that can flourish in soils with high saline content or under drought conditions. In particular, increased salt tolerance during the germination stage of a plant enhances survival and yield. Presently disclosed transcription factor genes, including G22, G196, G226, G312, G482, G801, G867, G922, G1836, G2110, and their equivalogs that provide increased salt tolerance during germination, the seedling stage, and throughout a plant's life cycle, would find particular value for imparting survival and yield in areas where a particular crop would not normally prosper.

Nutrient uptake and utilization: nitrogen and phosphorus. Presently disclosed transcription factor genes introduced into plants provide a means to improve uptake of essential nutrients, including nitrogenous compounds, phosphates, potassium, and trace minerals. The enhanced performance of, for example, G225, G226, G839, G1792, and other overexpressing lines under low nitrogen, and G545, G561, G911, G1946 under low phosphorous conditions indicate that these genes and their equivalogs can be used to engineer crops that could thrive under conditions of reduced nutrient availability. Phosphorus, in particular, tends to be a limiting nutrient in soils and is generally added as a component in fertilizers. Young plants have a rapid intake of phosphate and sufficient phosphate is important for yield of root crops such as carrot, potato and parsnip.

The effect of these modifications is to increase the seedling germination and range of ornamental and crop plants. The utilities of presently disclosed transcription factor genes conferring tolerance to conditions of low nutrients also include cost savings to the grower by reducing the amounts of fertilizer needed, environmental benefits of reduced fertilizer runoff into watersheds; and improved yield and stress tolerance. In addition, by providing improved nitrogen uptake capability, these genes can be used to alter seed protein amounts and/or composition in such a way that could impact yield as well as the nutritional value and production of various food products.

A number of the transcription factor-overexpressing lines make less anthocyanin on high sucrose plus glutamine indicates that these genes can be used to modify carbon and nitrogen status, and hence assimilate partitioning (assimilate partitioning refers to the manner in which an essential element, such as nitrogen, is distributed among different pools inside a plant, generally in a reduced form, for the purpose of transport to various tissues).

Increased tolerance of plants to oxidative stress. In plants, as in all living things, abiotic and biotic stresses induce the formation of oxygen radicals, including superoxide and peroxide radicals. This has the effect of accelerating senescence, particularly in leaves, with the resulting loss of yield and adverse effect on appearance. Generally, plants that have the highest level of defense mechanisms, such as, for example, polyunsaturated moieties of membrane lipids, are most likely to thrive under conditions that introduce oxidative stress (e.g., high light, ozone, water deficit, particularly in combination). Introduction of the presently disclosed transcription factor genes, including G477 and its equivalogs, that increase the level of oxidative stress defense mechanisms would provide beneficial effects on the yield and appearance of plants. One specific oxidizing agent, ozone, has been shown to cause significant foliar injury, which impacts yield and appearance of crop and ornamental plants. In addition to reduced foliar injury that would be found in ozone resistant plant created by transforming plants with some of the presently disclosed transcription factor genes, the latter have also been shown to have increased chlorophyll fluorescence (Yu-Sen Chang et al. (2001) *Bot. Bull. Acad. Sin.* 42: 265-272).

Decreased herbicide sensitivity. Presently disclosed transcription factor genes, including G343, G2133, G2517 and their equivalogs, that confer resistance or tolerance to herbicides (e.g., glyphosate) will find use in providing means to increase herbicide applications without detriment to desirable plants. This would allow for the increased use of a particular herbicide in a local environment, with the effect of increased detriment to undesirable species and less harm to transgenic, desirable cultivars.

Knockouts of a number of the presently disclosed transcription factor genes have been shown to be lethal to developing embryos. Thus, these genes are potentially useful as herbicide targets.

Hormone sensitivity. ABA plays regulatory roles in a host of physiological processes in all higher as well as in lower plants (Davies et al. (1991) Abscisic Acid: Physiology and Biochemistry. Bios Scientific Publishers, Oxford, UK; Zeevaart et al. (1988) *Ann Rev Plant Physiol. Plant Mol. Biol.* 49: 439-473; Shimizu-Sato et al. (2001) *Plant Physiol* 127: 1405-1413). ABA mediates stress tolerance responses in higher plants, is a key signal compound that regulates stomatal aperture and, in concert with other plant signaling compounds, is implicated in mediating responses to pathogens and wounding or oxidative damage (for example, see Larkindale et al. (2002) *Plant Physiol.* 128: 682-695). In seeds, ABA promotes seed development, embryo maturation, synthesis of storage products (proteins and lipids), desiccation tolerance, and is involved in maintenance of dormancy (inhibition of germination), and apoptosis (Zeevaart et al. (1988) *Ann Rev Plant Physiol. Plant Mol. Biol.* 49: 439-473; Davies (1991), *supra*; Thomas (1993) *Plant Cell* 5: 1401-1410; and Bethke et al. (1999) *Plant Cell* 11: 1033-1046). ABA also affects plant architecture, including root growth and morphology and root-to-shoot ratios. ABA action and metabolism is modulated not only by environmental signals but also by endogenous signals generated by metabolic feedback, transport, hormonal cross-talk and developmental stage. Manipulation of ABA levels, and hence by extension the sensitivity to ABA, has been described as a very promising means to improve productivity, performance and architecture in plants Zeevaart (1999) in: Biochemistry and Molecular Biology of Plant Hormones, Hooykaas et al. eds, Elsevier Science pp 189-207; and Cutler et al. (1999) *Trends Plant Sci.* 4: 472-478).

A number of the presently disclosed transcription factor genes affect plant abscisic acid (ABA) sensitivity, including G546, G926, 1069, G1357, G1452, G1820, G2140, G2789. Thus, by affecting ABA sensitivity, these introduced transcription factor genes and their equivalogs would affect cold, drought, oxidative and other stress sensitivities, plant architecture, and yield.

Several other of the present transcription factor genes have been used to manipulate ethylene signal transduction and response pathways. These genes can thus be used to manipulate the processes influenced by ethylene, such as seed germination or fruit ripening, and to improve seed or fruit quality.

Diseases, pathogens and pests. A number of the presently disclosed transcription factor genes have been shown to or are likely to affect a plants response to various plant diseases, pathogens and pests. The offending organisms include fungal *pathogens Fusarium oxysporum, Botrytis cinerea,*

Sclerotinia sclerotiorum, and *Erysiphe orontii.* Bacterial pathogens to which resistance may be conferred include *Pseudomonas syringae.* Other problem organisms may potentially include nematodes, mollicutes, parasites, or herbivorous arthropods. In each case, one or more transformed transcription factor genes may provide some benefit to the plant to help prevent or overcome infestation, or be used to manipulate any of the various plant responses to disease. These mechanisms by which the transcription factors work could include increasing surface waxes or oils, surface thickness, or the activation of signal transduction pathways that regulate plant defense in response to attacks by herbivorous pests (including, for example, protease inhibitors). Another means to combat fungal and other pathogens is by accelerating local cell death or senescence, mechanisms used to impair the spread of pathogenic microorganisms throughout a plant. For instance, the best known example of accelerated cell death is the resistance gene-mediated hypersensitive response, which causes localized cell death at an infection site and initiates a systemic defense response. . Because many defenses, signaling molecules, and signal transduction pathways are common to defense against different pathogens and pests, such as fungal, bacterial, oomycete, nematode, and insect, transcription factors that are implicated in defense responses against the fungal pathogens tested may also function in defense against other pathogens and pests. These transcription factors include, for example, G28, G1792, G1880, G1919, G1950 (improved resistance or tolerance to *Botrytis*), G1047, G1792 (improved resistance or tolerance to *Fusarium*), G19, G28, G409, G1266, G1363, G1792 (improved resistance or tolerance to *Erysiphe*), G545 (improved resistance or tolerance to *Pseudomonas*), G28, G1927 (improved resistance or tolerance to *Sclerotinia*), and their equivalogs.

Growth regulator: sugar sensing. In addition to their important role as an energy source and structural component of the plant cell, sugars are central regulatory molecules that control several aspects of plant physiology, metabolism and development (Hsieh et al. (1998) *Proc. Natl. Acad. Sci.* 95: 13965-13970). It is thought that this control is achieved by regulating gene expression and, in higher plants, sugars have been shown to repress or activate plant genes involved in many essential processes such as photosynthesis, glyoxylate metabolism, respiration, starch and sucrose synthesis and degradation, pathogen response, wounding response, cell cycle regulation, pigmentation, flowering and senescence. The mechanisms by which sugars control gene expression are not understood.

Because sugars are important signaling molecules, the ability to control either the concentration of a signaling sugar or how the plant perceives or responds to a signaling sugar could be used to control plant development, physiology or metabolism. For example, the flux of sucrose (a disaccharide sugar used for systemically transporting carbon and energy in most plants) has been shown to affect gene expression and alter storage compound accumulation in seeds. Manipulation of the sucrose signaling pathway in seeds may therefore cause seeds to have more protein, oil or carbohydrate, depending on the type of manipulation. Similarly, in tubers, sucrose is converted to starch which is used as an energy store.

It is thought that sugar signaling pathways may partially determine the levels of starch synthesized in the tubers. The manipulation of sugar signaling in tubers could lead to tubers with a higher starch content.

Thus, the presently disclosed transcription factor genes that manipulate the sugar signal transduction pathway, including G241, G254, G567, G680, G912, G1804, G481, G867, G1225, along with their equivalogs, may lead to altered gene expression to produce plants with desirable traits. In particular, manipulation of sugar signal transduction pathways could be used to alter source-sink relationships in seeds, tubers, roots and other storage organs leading to increase in yield.

Growth regulator: C/N sensing. Nitrogen and carbon metabolism are tightly linked in almost every biochemical pathway in the plant. Carbon metabolites regulate genes involved in N acquisition and metabolism, and are known to affect germination and the expression of photosynthetic genes (Coruzzi et al. (2001) *Plant Physiol.* 125: 61-64) and hence growth. Early studies on nitrate reductase (NR) in 1976 showed that NR activity could be affected by Glc/Suc (Crawford (1995) *Plant Cell* 7: 859-886; Daniel-Vedele et al. (1996*) CR Acad Sci Paris* 319: 961-968). Those observations were supported by later experiments that showed sugars induce NR mRNA in dark-adapted, green seedlings (Cheng CL, et al. (1992) *Proc Natl Acad Sci* USA 89: 1861-1864). C and N may have antagonistic relationships as signaling molecules; light induction of NR activity and mRNA levels can be mimicked by C metabolites and N-metabolites cause repression of NR induction in tobacco (Vincentz et al. (1992) *Plant J* 3: 315-324). Gene regulation by C/N status has been demonstrated for a number of N-metabolic genes (Stitt (1999) *Curr. Opin. Plant. Biol.* 2: 178-186); Coruzzi et al. (2001) *supra*). Thus, transcription factor genes that affect C/N sensing, such as G1816, can be used to alter or improve germination and growth under nitrogen-limiting conditions.

Flowering time: early and late flowering. Presently disclosed transcription factor genes that accelerate flowering, which include G157, G180, G183, G485, G490, G590, G789, G1225, G1494, G1820, G1841, G1842, G1843, G1946, G2010, G2144, G2295, G2347, G2509, and their functional equivalogs, could have valuable applications in such programs, since they allow much faster generation times. In a number of species, for example, broccoli, cauliflower, where the reproductive parts of the plants constitute the crop and the vegetative tissues are discarded, it would be advantageous to accelerate time to flowering. Accelerating flowering could shorten crop and tree breeding programs. Additionally, in some instances, a faster generation time would allow additional harvests of a crop to be made within a given growing season. A number of *Arabidopsis* genes have already been shown to accelerate flowering when constitutively expressed. These include LEAFY, APETALA1 and CONSTANS (Mandel et al. (1995) *Nature* 377: 522-524; Weigel and Nilsson (1995) *Nature* 377:et al. 495-500; Simon et al. (1996) *Nature* 384: 59-62).

By regulating the expression of potential flowering using inducible promoters, flowering could be triggered by application of an inducer chemical. This would allow flowering to be synchronized across

a crop and facilitate more efficient harvesting. Such inducible systems could also be used to tune the flowering of crop varieties to different latitudes. At present, species such as soybean and cotton are available as a series of maturity groups that are suitable for different latitudes on the basis of their flowering time (which is governed by day-length). A system in which flowering could be chemically controlled would allow a single high-yielding northern maturity group to be grown at any latitude. In southern regions such plants could be grown for longer periods before flowering was induced, thereby increasing yields. In more northern areas, the induction would be used to ensure that the crop flowers prior to the first winter frosts.

In a sizeable number of species, for example, root crops, where the vegetative parts of the plants constitute the crop and the reproductive tissues are discarded, it is advantageous to identify and incorporate transcription factor genes that delay or prevent flowering in order to prevent resources being diverted into reproductive development. For example, G8, G47, G157, G192, G214, G231; G361, G362, G562, G736, G748, G859, G910, G913, G971, G1051, G1052, G1357, G1452, G1478, G1804, G1895, G1945, G2007, G2133, G2155, G2838 and equivalogs, delay flowering time in transgenic plants. Extending vegetative development with presently disclosed transcription factor genes could thus bring about large increases in yields. Prevention of flowering can help maximize vegetative yields and prevent escape of genetically modified organism (GMO) pollen.

Presently disclosed transcription factors that extend flowering time have utility in engineering plants with longer-lasting flowers for the horticulture industry, and for extending the time in which the plant is fertile.

A number of the presently disclosed transcription factors may extend flowering time, and delay flower abscission, which would have utility in engineering plants with longer-lasting flowers for the horticulture industry. This would provide a significant benefit to the ornamental industry, for both cut flowers and woody plant varieties (of, for example, maize), as well as have the potential to lengthen the fertile period of a plant, which could positively impact yield and breeding programs.

General development and morphology: flower structure and inflorescence: architecture, altered flower organs, reduced fertility, multiple alterations, aerial rosettes, branching, internode distance, terminal flowers and phase change. Presently disclosed transgenic transcription factors such as G353; G354, G638; G779; G988; G1063; G1075; G1140; G1449; G1499; G2143; G2557, G2838, G2839 and their equivalogs, may be used to create plants with larger flowers or arrangements of flowers that are distinct from wild-type or non-transformed cultivars. This would likely have the most value for the ornamental horticulture industry, where larger flowers or interesting floral configurations are generally preferred and command the highest prices.

Flower structure may have advantageous or deleterious effects on fertility, and could be used, for example, to decrease fertility by the absence, reduction or screening of reproductive components. In fact, plants that overexpress a sizable number of the presently disclosed transcription factor genes e.g., G470,

G779, G988, G1075, G1140, G1499, G1947, G2143, G2557 and their functional equivalogs, possess reduced fertility; flowers are infertile and fail to yield seed. These could be desirable traits, as low fertility could be exploited to prevent or minimize the escape of the pollen of genetically modified organisms (GMOs) into the environment.

The alterations in shoot architecture seen in the lines transformed with G47, G1063, G1645, G2143, and their functional equivalogs indicates that these genes and their equivalogs can be used to manipulate inflorescence branching patterns. This could influence yield and offer the potential for more effective harvesting techniques. For example, a "self pruning" mutation of tomato results in a determinate growth pattern and facilitates mechanical harvesting (Pnueli et al. (2001) *Plant Cell* 13(12): 2687-702).

One interesting application for manipulation of flower structure, for example, by introduced transcription factors could be in the increased production of edible flowers or flower parts, including saffron, which is derived from the stigmas of *Crocus sathus*.

Genes that later silique conformation in brassicates may be used to modify fruit ripening processes in brassicates and other plants, which may positively affect seed or fruit quality.

A number of the presently disclosed transcription factors may affect the timing of phase changes in plants. Since the timing or phase changes generally affects a plant's eventual size, these genes may prove beneficial by providing means for improving yield and biomass.

General development and morphology: shoot meristem and branching patterns. Several of the presently disclosed transcription factor genes, including G390 and G391, and G1794, when introduced into plants, have been shown to cause stem bifurcations in developing shoots in which the shoot meristems split to form two or three separate shoots. These transcription factors and their functional equivalogs may thus be used to manipulate branching. This would provide a unique appearance, which may be desirable in ornamental applications, and may be used to modify lateral branching for use in the forestry industry. A reduction in the formation of lateral branches could reduce knot formation. Conversely, increasing the number of lateral branches could provide utility when a plant is used as a view- or windscreen.

General development and morphology: apical dominance: The modified expression of presently disclosed transcription factors (e.g., G47, G211, G1255, G1275, G1411, G1488, G1794, G2509 and their equivalogs) that reduce apical dominance could be used in ornamental horticulture, for example, to modify plant architecture, for example, to produce a shorter, more bushy stature than wild type. The latter form would have ornamental utility as well as provide increased resistance to lodging.

General development and morphology: trichome density, development or structure. Several of the presently disclosed transcription factor genes have been used to modify trichome number, density, trichome cell fate, amount of trichome products produced by plants, or produce ectopic trichome

formation. These include G225; G226, G247; G362, G370; G585, G634, G676, G682, G1332, G1452, G1995, G2826, and G2838. In most cases where the metabolic pathways are impossible to engineer, increasing trichome density or size on leaves may be the only way to increase plant productivity. Thus, by increasing trichome density, size or type, these trichome-affecting genes and their functional equivalogs would have profound utilities in molecular farming practices by making use of trichomes as a manufacturing system for complex secondary metabolites.

Trichome glands on the surface of many higher plants produce and secrete exudates that give protection from the elements and pests such as insects, microbes and herbivores. These exudates may physically immobilize insects and spores, may be insecticidal or ant-microbial or they may act as allergens or irritants to protect against herbivores. By modifying trichome location, density or activity with presently disclosed transcription factors that modify these plant characteristics, plants that are better protected and higher yielding may be the result.

A potential application for these trichome-affecting genes and their equivalogs also exists in cotton: cotton fibers are modified unicellular trichomes that develop from the outer ovule epidermis. In fact, only about 30% of these epidermal cells develop into trichomes, but all have the potential to develop a trichome fate. Trichome-affecting genes can trigger an increased number of these cells to develop as trichomes and thereby increase the yield of cotton fibers. Since the mallow family is closely related to the Brassica family, genes involved in trichome formation will likely have homologs in cotton or function in cotton.

If the effects on trichome patterning reflect a general change in heterochronic processes, trichome-affecting transcription factors or their equivalogs can be used to modify the way meristems and/or cells develop during different phases of the plant life cycle. In particular, altering the timing of phase changes could afford positive effects on yield and biomass production.

General development and morphology: stem morphology and altered vascular tissue structure. Plants transformed with transcription factor genes that modify stem morphology or lignin content may be used to affect overall plant architecture and the distribution of lignified fiber cells within the stem.

Modulating lignin content might allow the quality of wood used for furniture or construction to be improved. Lignin is energy rich; increasing lignin composition could therefore be valuable in raising the energy content of wood used for fuel. Conversely, the pulp and paper industries seek wood with a reduced lignin content. Currently, lignin must be removed in a costly process that involves the use of many polluting chemicals. Consequently, lignin is a serious barrier to efficient pulp and paper production (Tzfira et al. (1998) *TIBTECH* 16: 439-446; Robinson (1999) *Nature Biotechnology* 17: 27-30). In addition to forest biotechnology applications, changing lignin content by selectively expressing or repressing transcription factors in fruits and vegetables might increase their palatability.

Transcription factors that modify stem structure, including G47, G438, G748, G988, G1488 and their equivalogs, may also be used to achieve reduction of higher-order shoot development, resulting in

significant plant architecture modification. Overexpression of the genes that encode these transcription factors in woody plants might result in trees that lack side branches, and have fewer knots in the wood. Altering branching patterns could also have applications amongst ornamental and agricultural crops. For example, applications might exist in any species where secondary shoots currently have to be removed manually, or where changes in branching pattern could increase yield or facilitate more efficient harvesting.

General development and morphology: altered root development. By modifying the structure or development of roots by transforming into a plant one or more of the presently disclosed transcription factor genes, including G225, G226, G1482, and their equivalogs, plants may be produced that have the capacity to thrive in otherwise unproductive soils. For example, grape roots extending further into rocky soils would provide greater anchorage, greater coverage with increased branching, or would remain viable in waterlogged soils, thus increasing the effective planting range of the crop and/or increasing yield and survival. It may be advantageous to manipulate a plant to produce short roots, as when a soil in which the plant will be growing is occasionally flooded, or when pathogenic fungi or disease-causing nematodes are prevalent.

General development and morphology: seed development, ripening and germination rate. A number of the presently disclosed transcription factor genes (e.g., G979) have been shown to modify seed development and germination rate, including when the seeds are in conditions normally unfavorable for germination (e.g., cold, heat or salt stress, or in the presence of ABA), and may, along with functional equivalogs, thus be used to modify and improve germination rates under adverse conditions.

General development and morphology: cell differentiation and cell proliferation. Several of the disclosed transcription factors regulate cell proliferation and/or differentiation, including G1540 and its functional equivalogs. Control of these processes could have valuable applications in plant transformation, cell culture or micro-propagation systems, as well as in control of the proliferation of particular useful tissues or cell types. Transcription factors that induce the proliferation of undifferentiated cells can be operably linked with an inducible promoter to promote the formation of callus that can be used for transformation or production of cell suspension cultures. Transcription factors that prevent cells from differentiating, such as G1540 or its equivalogs, could be used to confer stem cell identity to cultured cells. Transcription factors that promote differentiation of shoots could be used in transformation or micro-propagation systems, where regeneration of shoots from callus is currently problematic. In addition, transcription factors that regulate the differentiation of specific tissues could be used to increase the proportion of these tissues in a plant. Genes that promote the differentiation of carpel tissue could be introduced into commercial species to induce formation of increased numbers of carpels or fruits. A particular application might exist in saffron, one of the world's most expensive spices.

Saffron filaments, or threads, are actually the dried stigmas of the saffron flower, *Crocus sativus Linneaus*. Each flower contains only three stigmas, and more than 75,000 of these flowers are needed to produce just one pound of saffron filaments. An increase in carpel number would increase the quantity of stigmatic tissue and improve yield.

General development and morphology: cell expansion. Plant growth results from a combination of cell division and cell expansion. Transcription factors may be useful in regulation of cell expansion. Altered regulation of cell expansion could affect stem length, an important agronomic characteristic. For instance, short cultivars of wheat contributed to the Green Revolution, because plants that put fewer resources into stem elongation allocate more resources into developing seed and produce higher yield. These plants are also less vulnerable to wind and rain damage. These cultivars were found to be altered in their sensitivity to gibberellins, hormones that regulate stem elongation through control of both cell expansion and cell division. Altered cell expansion in leaves could also produce novel and ornamental plant forms.

General development and morphology: phase change and floral reversion. Transcription factors that regulate phase change can modulate the developmental programs of plants and regulate developmental plasticity of the shoot meristem. In particular, these genes might be used to manipulate seasonality and influence whether plants display an annual or perennial habit.

General development and morphology: rapid development. A number of the presently disclosed transcription factor genes, including G2430, have been shown to have significant effects on plant growth rate and development. These observations have included, for example, more rapid or delayed growth and development of reproductive organs. Thus, by causing more rapid development, G2430 and its functional equivalogs would prove useful for regions with short growing seasons; other transcription factors that delay development may be useful for regions with longer growing seasons. Accelerating plant growth would also improve early yield or increase biomass at an earlier stage, when such is desirable (for example, in producing forestry products or vegetable sprouts for consumption). Transcription factors that promote faster development such as G2430 and its functional equivalogs may also be used to modify the reproductive cycle of plants.

General development and morphology: slow growth rate. A number of the presently disclosed transcription factor genes, including G652 and G1335, have been shown to have significant effects on retarding plant growth rate and development. These observations have included, for example, delayed growth and development of reproductive organs. Slow growing plants may be highly desirable to ornamental horticulturists, both for providing house plants that display little change in their appearance over time, or outdoor plants for which wild-type or rapid growth is undesirable (e.g., ornamental palm

trees). Slow growth may also provide for a prolonged fruiting period, thus extending the harvesting season, particularly in regions with long growing seasons. Slow growth could also provide a prolonged period in which pollen is available for improved self- or cross-fertilization, or cross-fertilization of cultivars that normally flower over non-overlapping time periods. The latter aspect may be particularly useful to plants comprising two or more distinct grafted cultivars (e.g., fruit trees) with normally non-overlapping flowering periods.

General development and morphology: senescence. Presently disclosed transcription factor genes may be used to alter senescence responses in plants. Although leaf senescence is thought to be an evolutionary adaptation to recycle nutrients, the ability to control senescence in an agricultural setting has significant value. For example, a delay in leaf senescence in some maize hybrids is associated with a significant increase in yields and a delay of a few days in the senescence of soybean plants can have a large impact on yield. In an experimental setting, tobacco plants engineered to inhibit leaf senescence had a longer photosynthetic lifespan, and produced a 50% increase in dry weight and seed yield (Gan and Amasino (1995) *Science* 270: 1986-1988). Delayed flower senescence caused by overexpression of transcription factors may generate plants that retain their blossoms longer and this may be of potential interest to the ornamental horticulture industry, and delayed foliar and fruit senescence could improve post-harvest shelf-life of produce.

Premature senescence caused by, for example, G636, G1463, G1944 and their equivalogs may be used to improve a plant's response to disease and hasten fruit ripening.

Growth rate and development: lethality and necrosis. Overexpression of transcription factors, for example, G12, G24, G877, G1519 and their equivalogs that have a role in regulating cell death may be used to induce lethality in specific tissues or necrosis in response to pathogen attack. For example, if a transcription factor gene inducing lethality or necrosis was specifically active in gametes or reproductive organs, its expression in these tissues would lead to ablation and subsequent male or female sterility. Alternatively, under pathogen-regulated expression, a necrosis-inducing transcription factor can restrict the spread of a pathogen infection through a plant.

Plant size: large plants. Plants overexpressing G1073 and G1451, for example, have been shown to be larger than controls. For some ornamental plants, the ability to provide larger varieties with these genes or their equivalogs may be highly desirable. For many plants, including fruit-bearing trees, trees that are used for lumber production, or trees and shrubs that serve as view or wind screens, increased stature provides improved benefits in the forms of greater yield or improved screening. Crop species may also produce higher yields on larger cultivars, particularly those in which the vegetative portion of the plant is edible.

Plant size: large seedlings. Presently disclosed transcription factor genes, that produce large seedlings can be used to produce crops that become established faster. Large seedlings are generally hardier, less vulnerable to stress, and better able to out-compete weed species. Seedlings transformed with presently disclosed transcription factors, including G2346 and G2838, for example, have been shown to possess larger cotyledons and were more developmentally advanced than control plants. Rapid seedling development made possible by manipulating expression of these genes or their equivalogs is likely to reduce loss due to diseases particularly prevalent at the seedling stage (e.g., damping off) and is thus important for survivability of plants germinating in the field or in controlled environments.

Plant size: dwarfed plants. Presently disclosed transcription factor genes, including G24; G343, G353, G354, G362, G370; G1008, G1277, G1543, G1794, G1958 and their equivalogs, for example, that can be used to decrease plant stature are likely to produce plants that are more resistant to damage by wind and rain, have improved lodging resistance, or more resistant to heat or low humidity or water deficit. Dwarf plants are also of significant interest to the ornamental horticulture industry, and particularly for home garden applications for which space availability may be limited.

Plant size: fruit size and number. Introduction of presently disclosed transcription factor genes that affect fruit size will have desirable impacts on fruit size and number, which may comprise increases in yield for fruit crops, or reduced fruit yield, such as when vegetative growth is preferred (e.g., with bushy ornamentals, or where fruit is undesirable, as with ornamental olive trees).

Leaf morphology: dark leaves. Color-affecting components in leaves include chlorophylls (generally green), anthocyanins (generally red to blue) and carotenoids (generally yellow to red). Transcription factor genes that increase these pigments in leaves, including G674, G912, G1063, G1357, G1452, G1482, G1499, G1792, G1863, G1888, G2143, G2557, G2838 and their equivalogs, may positively affect a plant's value to the ornamental horticulture industry. Variegated varieties, in particular, would show improved contrast. Other uses that result from overexpression of transcription factor genes include improvements in the nutritional value of foodstuffs. For example, lutein is an important nutraceutical; lutein-rich diets have been shown to help prevent age-related macular degeneration (ARMD), the leading cause of blindness in elderly people. Consumption of dark green leafy vegetables has been shown in clinical studies to reduce the risk of ARMD.

Enhanced chlorophyll and carotenoid levels could also improve yield in crop plants. Lutein, like other xanthophylls such as zeaxanthin and violaxanthin, is an essential component in the protection of the plant against the damaging effects of excessive light. Specifically, lutein contributes, directly or indirectly, to the rapid rise of non-photochemical quenching in plants exposed to high light. Crop plants engineered to contain higher levels of lutein could therefore have improved photo-protection, leading to less oxidative damage and better growth under high light (e.g., during long summer days, or at higher

altitudes or lower latitudes than those at which a non-transformed plant would survive). Additionally, elevated chlorophyll levels increases photosynthetic capacity.

Leaf morphology: changes in leaf shape. Presently disclosed transcription factors produce marked and diverse effects on leaf development and shape. The transcription factors include G211, G353, G674, G736, G1063, G1146, G1357, G1452, G1494, G1543, G1863, G2143, G2144, and their equivalogs. At early stages of growth, transgenic seedlings have developed narrow, upward pointing leaves with long petioles, possibly indicating a disruption in circadian-clock controlled processes or nyctinastic movements. Other transcription factor genes can be used to alter leaf shape in a significant manner from wild type, some of which may find use in ornamental applications.

Leaf morphology: altered leaf size. Large leaves, such as those produced in plants overexpressing G189, G1451, G2430 and their functional equivalogs, generally increase plant biomass. This provides benefit for crops where the vegetative portion of the plant is the marketable portion.

Leaf morphology: light green and variegated leaves. Transcription factor genes such as G635, G1494, G2144 and their equivalogs that provide an altered appearance may positively affect a plant's value to the ornamental horticulture industry.

Leaf morphology: glossy leaves. Transcription factor genes such as G30, G1792, G2583 and their equivalogs that induce the formation of glossy leaves generally do so by elevating levels of epidermal wax. Thus, the genes could be used to engineer changes in the composition and amount of leaf surface components, including waxes. The ability to manipulate wax composition, amount, or distribution could modify plant tolerance to drought and low humidity, or resistance to insects or pathogens. Additionally, wax may be a valuable commodity in some species, and altering its accumulation and/or composition could enhance yield.

Seed morphology: altered seed coloration. Presently disclosed transcription factor genes, including G156, G2105, G2085 have also been used to modify seed color, which, along with the equivalogs of these genes, could provide added appeal to seeds or seed products.

Seed morphology: altered seed size and shape. The introduction of presently disclosed transcription factor genes into plants that increase (e.g., G450; G584; G1255; G2085; G2105; G2114) or decrease (e.g., G1040). the size of seeds may have a significant impact on yield and appearance, particularly when the product is the seed itself (e.g., in the case of grains, legumes, nuts, etc.). Seed size, in addition to seed coat integrity, thickness and permeability, seed water content and a number of other

components including antioxidants and oligosaccharides, also affects affect seed longevity in storage, with larger seeds often being more desirable for prolonged storage.

Transcription factor genes that alter seed shape, including G1040, G1062, G1255 and their equivalogs may have both ornamental applications and improve or broaden the appeal of seed products.

Leaf biochemistry: increased leaf wax . Overexpression of transcription factors genes, including G975, G1792 and G2085 and their equivalogs, which results in increased leaf wax could be used to manipulate wax composition, amount, or distribution. These transcription factors can improve yield in those plants and crops from which wax is a valuable product. The genes may also be used to modify plant tolerance to drought and/or low humidity or resistance to insects, as well as plant appearance (glossy leaves). The effect of increased wax deposition on leaves of a plant like may improve water use efficiency. Manipulation of these genes may reduce the wax coating on sunflower seeds; this wax fouls the oil extraction system during sunflower seed processing for oil. For the latter purpose or any other where wax reduction is valuable, antisense or co-suppression of the transcription factor genes in a tissue-specific manner would be valuable.

Leaf biochemistry: leaf prenyl lipids, including tocopherol. Prenyl lipids play a role in anchoring proteins in membranes or membranous organelles. Thus modifying the prenyl lipid content of seeds and leaves could affect membrane integrity and function. One important group of prenyl lipids, the tocopherols, have both anti-oxidant and vitamin E activity. A number of presently disclosed transcription factor genes, including G214, G652, G748, G987, G1543, and G2509, have been shown to modify the tocopherol composition of leaves in plants, and these genes and their equivalogs may thus be used to alter prenyl lipid content of leaves.

Leaf biochemistry: increased leaf insoluble sugars. Overexpression of a number of presently disclosed transcription factors, including G211, resulted in plants with altered leaf insoluble sugar content. This transcription factor and its equivalogs that alter plant cell wall composition have several potential applications including altering food digestibility, plant tensile strength, wood quality, pathogen resistance and in pulp production. In particular, hemicellulose is not desirable in paper pulps because of its lack of strength compared with cellulose. Thus modulating the amounts of cellulose vs. hemicellulose in the plant cell wall is desirable for the paper/lumber industry. Increasing the insoluble carbohydrate content in various fruits, vegetables, and other edible consumer products will result in enhanced fiber content. Increased fiber content would not only provide health benefits in food products, but might also increase digestibility of forage crops. In addition, the hemicellulose and pectin content of fruits and berries affects the quality of jam and catsup made from them. Changes in hemicellulose and pectin content could result in a superior consumer product.

Leaf biochemistry: increased leaf anthocyanin. Several presently disclosed transcription factor genes may be used to alter anthocyanin production in numerous plant species. Expression of presently disclosed transcription factor genes that increase flavonoid production in plants, including anthocyanins and condensed tannins, may be used to alter in pigment production for horticultural purposes, and possibly increasing stress resistance. G362, G663, G1482 and G1888 or their equivalogs, for example, could be used to alter anthocyanin production or accumulation. A number of flavonoids have been shown to have antimicrobial activity and could be used to engineer pathogen resistance. Several flavonoid compounds have health promoting effects such as inhibition of tumor growth, prevention of bone loss and prevention of the oxidation of lipids. Increased levels of condensed tannins, in forage legumes would be an important agronomic trait because they prevent pasture bloat by collapsing protein foams within the rumen. For a review on the utilities of flavonoids and their derivatives, refer to Dixon et al. (1999) *Trends Plant Sci.* 4: 394-400.

Leaf and seed biochemistry: altered fatty acid content. A number of the presently disclosed transcription factor genes have been shown to alter the fatty acid composition in plants, and seeds and leaves in particular. This modification suggests several utilities, including improving the nutritional value of seeds or whole plants. Dietary fatty acids ratios have been shown to have an effect on, for example, bone integrity and remodeling (see, for example, Weiler (2000) *Pediatr. Res.* 47:5 692-697). The ratio of dietary fatty acids may alter the precursor pools of long-chain polyunsaturated fatty acids that serve as precursors for prostaglandin synthesis. In mammalian connective tissue, prostaglandins serve as important signals regulating the balance between resorption and formation in bone and cartilage. Thus dietary fatty acid ratios altered in seeds may affect the etiology and outcome of bone loss.

Transcription factors that reduce leaf fatty acids, for example, 16:3 fatty acids, may be used to control thylakoid membrane development, including proplastid to chloroplast development. The genes that encode these transcription factors might thus be useful for controlling the transition from proplastid to chromoplast in fruits and vegetables. It may also be desirable to change the expression of these genes to prevent cotyledon greening in *Brassica napus* or *B. campestris* to avoid green oil due to early frost.

A number of transcription factor genes are involved in mediating an aspect of the regulatory response to temperature. These genes may be used to alter the expression of desaturases that lead to production of 18:3 and 16:3 fatty acids, the balance of which affects membrane fluidity and mitigates damage to cell membranes and photosynthetic structures at high and low temperatures.

Seed biochemistry: modified seed oil and fatty acid content. The composition of seeds, particularly with respect to seed oil amounts and/or composition, is very important for the nutritional and caloric value and production of various food and feed products. Several of the presently disclosed transcription factor genes in seed lipid saturation that alter seed oil content could be used to improve the heat stability of oils or to improve the nutritional quality of seed oil, by, for example, reducing the

number of calories in seed by decreasing oil or fatty acid content (e.g., G180; G192; G241; G1229; G1323; G1543) , increasing the number of calories in animal feeds by increasing oil or fatty acid content (e.g. G162; G291; G427; G590; G598; G629, G715; G849; G1198, G1471; G1526; G1640; G1646, G1750; G1777; G1793; G1838; G1902; G1946; G1948; G2123; G2138; G2830), altering seed oil content (G504; G509; G519; G561; G567; G892; G961; G974; G1143; G1226; G1451; G1478; G1496; G1672; G1677; G1765; G2509; G2343), or altering the ratio of saturated to unsaturated lipids comprising the oils (e.g. G869; G1417; G2192).

Seed biochemistry: modified seed protein content. As with seed oils, the composition of seeds, particularly with respect to protein amounts and/or composition, is very important for the nutritional value and production of various food and feed products. A number of the presently disclosed transcription factor genes modify the protein concentrations in seeds, including G162; G226; G1323; G1419; G1818, which increase seed protein, G427; G1777; G1903; G1946, which decrease seed protein, and G162; G241; G509; G567; G597; G849; G892; G988; G1478; G1634; G1637; G1652; G1677; G1820; G1958; G2509; G2117; G2509 , which alter seed protein content, would provide nutritional benefits, and may be used to prolong storage, increase seed pest or disease resistance, or modify germination rates.

Seed biochemistry: seed prenyl lipids. Prenyl lipids play a role in anchoring proteins in membranes or membranous organelles. Thus, modifying the prenyl lipid content of seeds and leaves could affect membrane integrity and function. A number of presently disclosed transcription factor genes have been shown to modify the tocopherol composition of plants. α-Tocopherol is better known as vitamin E. Tocopherols such as α- and γ-tocopherol both have anti-oxidant activity.

Seed biochemistry: seed glucosinolates. A number of glucosinolates have been shown to have anti-cancer activity; thus, increasing the levels or composition of these compounds by introducing several of the presently disclosed transcription factors, including G484 and G2340, can have a beneficial effect on human diet.

Glucosinolates are undesirable components of the oilseeds used in animal feed since they produce toxic effects. Low-glucosinolate varieties of canola, for example, have been developed to combat this problem. Glucosinolates form part of a plant's natural defense against insects. Modification of glucosinolate composition or quantity by introducing transcription factors that affect these characteristics can therefore afford increased protection from herbivores. Furthermore, in edible crops, tissue specific promoters can be used to ensure that these compounds accumulate specifically in tissues, such as the epidermis, which are not taken for consumption.

Seed biochemistry: increased seed anthocyanin. Several presently disclosed transcription factor genes may be used to alter anthocyanin production in the seeds of plants. As with leaf anthocyanins, expression of presently disclosed transcription factor genes that increase flavonoid (anthocyanins and condensed tannins) production in seeds, including G663 and its equivalogs, may be used to alter in pigment production for horticultural purposes, and possibly increasing stress resistance, antimicrobial activity and health promoting effects such as inhibition of tumor growth, prevention of bone loss and prevention of the oxidation of lipids.

Leaf and seed biochemistry: production of seed and leaf phytosterols: Presently disclosed transcription factor genes that modify levels of phytosterols in plants may have at least two utilities. First, phytosterols are an important source of precursors for the manufacture of human steroid hormones. Thus, regulation of transcription factor expression or activity could lead to elevated levels of important human steroid precursors for steroid semi-synthesis. For example, transcription factors that cause elevated levels of campesterol in leaves, or sitosterols and stigmasterols in seed crops, would be useful for this purpose. Phytosterols and their hydrogenated derivatives phytostanols also have proven cholesterol-lowering properties, and transcription factor genes that modify the expression of these compounds in plants would thus provide health benefits.

Root biochemistry: increased root anthocyanin. Presently disclosed transcription factor genes, including G663, may be used to alter anthocyanin production in the root of plants. As described above for seed anthocyanins, expression of presently disclosed transcription factor genes that increase flavonoid (anthocyanins and condensed tannins) production in seeds, including G663 and its equivalogs, may be used to alter in pigment production for horticultural purposes, and possibly increasing stress resistance, antimicrobial activity and health promoting effects such as inhibition of tumor growth, prevention of bone loss and prevention of the oxidation of lipids.

Light response/shade avoidance: altered cotyledon, hypocotyl, petiole development, altered leaf orientation, constitutive photomorphogenesis, photomorphogenesis in low light. Presently disclosed transcription factor genes, including G183; G354; G1322; G1331; G1488; G1494; G1794; G2144; and G2555, that modify a plant's response to light may be useful for modifying plant growth or development, for example, photomorphogenesis in poor light, or accelerating flowering time in response to various light intensities, quality or duration to which a non-transformed plant would not similarly respond. Examples of such responses that have been demonstrated include leaf number and arrangement, and early flower bud appearances. Elimination of shading responses may lead to increased planting densities with subsequent yield enhancement. As these genes may also alter plant architecture, they may find use in the ornamental horticulture industry.

Pigment; increased anthocyanin level in various plant organs and tissues. In addition to seed, leaves and roots, as mentioned above, several presently disclosed transcription factor genes can be used to alter anthocyanin levels in one or more tissues. The potential utilities of these genes include alterations in pigment production for horticultural purposes, and possibly increasing stress resistance, antimicrobial activity and health promoting effects such as inhibition of tumor growth, prevention of bone loss and prevention of the oxidation of lipids.

Miscellaneous biochemistry; diterpenes in leaves and other plant parts. Depending on the plant species, varying amounts of diverse secondary biochemicals (often lipophilic terpenes) are produced and exuded or volatilized by trichomes. These exotic secondary biochemicals, which are relatively easy to extract because they are on the surface of the leaf, have been widely used in such products as flavors and aromas, drugs, pesticides and cosmetics. Thus, the overexpression of genes that are used to produce diterpenes in plants may be accomplished by introducing transcription factor genes that induce said overexpression. One class of secondary metabolites, the diterpenes, can effect several biological systems such as tumor progression, prostaglandin synthesis and tissue inflammation. In addition, diterpenes can act as insect pheromones, termite allomones, and can exhibit neurotoxic, cytotoxic and antimitotic activities. As a result of this functional diversity, diterpenes have been the target of research several pharmaceutical ventures. In most cases where the metabolic pathways are impossible to engineer, increasing trichome density or size on leaves may be the only way to increase plant productivity.

Miscellaneous biochemistry; production of miscellaneous secondary metabolites. Microarray data suggests that flux through the aromatic amino acid biosynthetic pathways and primary and secondary metabolite biosynthetic pathways are up-regulated. Presently disclosed transcription factors have been shown to be involved in regulating alkaloid biosynthesis, in part by up-regulating the enzymes indole-3-glycerol phosphatase and strictosidine synthase. Phenylalanine ammonia lyase, chalcone synthase and trans-cinnamate mono-oxygenase are also induced, and are involved in phenylpropenoid biosynthesis.

Antisense and Co-suppression

In addition to expression of the nucleic acids of the invention as gene replacement or plant phenotype modification nucleic acids, the nucleic acids are also useful for sense and anti-sense suppression of expression, e.g., to down-regulate expression of a nucleic acid of the invention, e.g., as a further mechanism for modulating plant phenotype. That is, the nucleic acids of the invention, or subsequences or anti-sense sequences thereof, can be used to block expression of naturally occurring homologous nucleic acids. A variety of sense and anti-sense technologies are known in the art, e.g., as set forth in Lichtenstein and Nellen (1997) *Antisense Technology: A Practical Approach* IRL Press at Oxford University Press, Oxford, U.K. Antisense regulation is also described in Crowley et al. (1985) *Cell* 43: 633-641; Rosenberg et al. (1985) *Nature* 313: 703-706; Preiss et al. (1985) *Nature* 313: 27-32;

Melton (1985) *Proc. Natl. Acad. Sci.* 82: 144-148; Izant and Weintraub (1985) *Science* 229: 345-352; and Kim and Wold (1985) *Cell* 42: 129-138. Additional methods for antisense regulation are known in the art. Antisense regulation has been used to reduce or inhibit expression of plant genes in, for example in European Patent Publication No. 271988. Antisense RNA may be used to reduce gene expression to produce a visible or biochemical phenotypic change in a plant (Smith et al. (1988) *Nature*, 334: 724-726; Smith et al. (1990) *Plant Mol. Biol.* 14: 369-379). In general, sense or anti-sense sequences are introduced into a cell, where they are optionally amplified, e.g., by transcription. Such sequences include both simple oligonucleotide sequences and catalytic sequences such as ribozymes.

For example, a reduction or elimination of expression (i.e., a "knock-out") of a transcription factor or transcription factor homolog polypeptide in a transgenic plant, e.g., to modify a plant trait, can be obtained by introducing an antisense construct corresponding to the polypeptide of interest as a cDNA. For antisense suppression, the transcription factor or homolog cDNA is arranged in reverse orientation (with respect to the coding sequence) relative to the promoter sequence in the expression vector. The introduced sequence need not be the full length cDNA or gene, and need not be identical to the cDNA or gene found in the plant type to be transformed. Typically, the antisense sequence need only be capable of hybridizing to the target gene or RNA of interest. Thus, where the introduced sequence is of shorter length, a higher degree of homology to the endogenous transcription factor sequence will be needed for effective antisense suppression. While antisense sequences of various lengths can be utilized, preferably, the introduced antisense sequence in the vector will be at least 30 nucleotides in length, and improved antisense suppression will typically be observed as the length of the antisense sequence increases. Preferably, the length of the antisense sequence in the vector will be greater than 100 nucleotides. Transcription of an antisense construct as described results in the production of RNA molecules that are the reverse complement of mRNA molecules transcribed from the endogenous transcription factor gene in the plant cell.

Suppression of endogenous transcription factor gene expression can also be achieved using a ribozyme. Ribozymes are RNA molecules that possess highly specific endoribonuclease activity. The production and use of ribozymes are disclosed in U.S. Patent No. 4,987,071 and U.S. Patent No. 5,543,508. Synthetic ribozyme sequences including antisense RNAs can be used to confer RNA cleaving activity on the antisense RNA, such that endogenous mRNA molecules that hybridize to the antisense RNA are cleaved, which in turn leads to an enhanced antisense inhibition of endogenous gene expression.

Vectors in which RNA encoded by a transcription factor or transcription factor homolog cDNA is over-expressed can also be used to obtain co-suppression of a corresponding endogenous gene, e.g., in the manner described in U.S. Patent No. 5,231,020 to Jorgensen. Such co-suppression (also termed sense suppression) does not require that the entire transcription factor cDNA be introduced into the plant cells, nor does it require that the introduced sequence be exactly identical to the endogenous transcription factor gene of interest. However, as with antisense suppression, the suppressive efficiency will be enhanced as specificity of hybridization is increased, e.g., as the introduced sequence is lengthened, and/or as the

sequence similarity between the introduced sequence and the endogenous transcription factor gene is increased.

Vectors expressing an untranslatable form of the transcription factor mRNA, e.g., sequences comprising one or more stop codon, or nonsense mutation) can also be used to suppress expression of an endogenous transcription factor, thereby reducing or eliminating its activity and modifying one or more traits. Methods for producing such constructs are described in U.S. Patent No. 5,583,021. Preferably, such constructs are made by introducing a premature stop codon into the transcription factor gene. Alternatively, a plant trait can be modified by gene silencing using double-strand RNA (Sharp (1999) *Genes and Development* 13: 139-141). Another method for abolishing the expression of a gene is by insertion mutagenesis using the T-DNA of *Agrobacterium tumefaciens*. After generating the insertion mutants, the mutants can be screened to identify those containing the insertion in a transcription factor or transcription factor homolog gene. Plants containing a single transgene insertion event at the desired gene can be crossed to generate homozygous plants for the mutation. Such methods are well known to those of skill in the art (See for example Koncz et al. (1992) *Methods in Arabidopsis Research,* World Scientific Publishing Co. Pte. Ltd., River Edge, NJ).

Alternatively, a plant phenotype can be altered by eliminating an endogenous gene, such as a transcription factor or transcription factor homolog, e.g., by homologous recombination (Kempin et al. (1997) *Nature* 389: 802-803).

A plant trait can also be modified by using the Cre-lox system (for example, as described in US Pat. No. 5,658,772). A plant genome can be modified to include first and second lox sites that are then contacted with a Cre recombinase. If the lox sites are in the same orientation, the intervening DNA sequence between the two sites is excised. If the lox sites are in the opposite orientation, the intervening sequence is inverted.

The polynucleotides and polypeptides of this invention can also be expressed in a plant in the absence of an expression cassette by manipulating the activity or expression level of the endogenous gene by other means, such as, for example, by ectopically expressing a gene by T-DNA activation tagging (Ichikawa et al. (1997) *Nature* 390 698-701; Kakimoto et al. (1996) *Science* 274: 982-985). This method entails transforming a plant with a gene tag containing multiple transcriptional enhancers and once the tag has inserted into the genome, expression of a flanking gene coding sequence becomes deregulated. In another example, the transcriptional machinery in a plant can be modified so as to increase transcription levels of a polynucleotide of the invention (*See*, e.g., PCT Publications WO 96/06166 and WO 98/53057 which describe the modification of the DNA-binding specificity of zinc finger proteins by changing particular amino acids in the DNA-binding motif).

The transgenic plant can also include the machinery necessary for expressing or altering the activity of a polypeptide encoded by an endogenous gene, for example, by altering the phosphorylation state of the polypeptide to maintain it in an activated state.

Transgenic plants (or plant cells, or plant explants, or plant tissues) incorporating the polynucleotides of the invention and/or expressing the polypeptides of the invention can be produced by a variety of well established techniques as described above. Following construction of a vector, most typically an expression cassette, including a polynucleotide, e.g., encoding a transcription factor or transcription factor homolog, of the invention, standard techniques can be used to introduce the polynucleotide into a plant, a plant cell, a plant explant or a plant tissue of interest. Optionally, the plant cell, explant or tissue can be regenerated to produce a transgenic plant.

The plant can be any higher plant, including gymnosperms, monocotyledonous and dicotyledenous plants. Suitable protocols are available for *Leguminosae* (alfalfa, soybean, clover, etc.), *Umbelliferae* (carrot, celery, parsnip), *Cruciferae* (cabbage, radish, rapeseed, broccoli, etc.), *Curcurbitaceae* (melons and cucumber), *Gramineae* (wheat, corn, rice, barley, millet, etc.), *Solanaceae* (potato, tomato, tobacco, peppers, etc.), and various other crops. See protocols described in Ammirato et al., Eds., (1984) Handbook of Plant Cell Culture –Crop Species, Macmillan Publ. Co., New York, NY; Shimamoto et al. (1989) *Nature* 338: 274-276; Fromm et al. (1990) *Bio/Technol.* 8: 833-839; and Vasil et al. (1990) *Bio/Technol.* 8: 429-434.

Transformation and regeneration of both monocotyledonous and dicotyledonous plant cells are now routine, and the selection of the most appropriate transformation technique will be determined by the practitioner. The choice of method will vary with the type of plant to be transformed; those skilled in the art will recognize the suitability of particular methods for given plant types. Suitable methods can include, but are not limited to: electroporation of plant protoplasts; liposome-mediated transformation; polyethylene glycol (PEG) mediated transformation; transformation using viruses; micro-injection of plant cells; micro-projectile bombardment of plant cells; vacuum infiltration; and *Agrobacterium tumefaciens* mediated transformation. Transformation means introducing a nucleotide sequence into a plant in a manner to cause stable or transient expression of the sequence.

Successful examples of the modification of plant characteristics by transformation with cloned sequences which serve to illustrate the current knowledge in this field of technology, and which are herein incorporated by reference, include: U.S. Patent Nos. 5,571,706; 5,677,175; 5,510,471; 5,750,386; 5,597,945; 5,589,615; 5,750,871; 5,268,526; 5,780,708; 5,538,880; 5,773,269; 5,736,369 and 5,610,042.

Following transformation, plants are preferably selected using a dominant selectable marker incorporated into the transformation vector. Typically, such a marker will confer antibiotic or herbicide resistance on the transformed plants, and selection of transformants can be accomplished by exposing the plants to appropriate concentrations of the antibiotic or herbicide.

After transformed plants are selected and grown to maturity, those plants showing a modified trait are identified. The modified trait can be any of those traits described above. Additionally, to confirm that the modified trait is due to changes in expression levels or activity of the polypeptide or polynucleotide of the invention can be determined by analyzing mRNA expression using Northern blots, RT-PCR or microarrays, or protein expression using immunoblots or Western blots or gel shift assays.

Integrated Systems – Sequence Identity

Additionally, the present invention may be an integrated system, computer or computer readable medium that comprises an instruction set for determining the identity of one or more sequences in a database. In addition, the instruction set can be used to generate or identify sequences that meet any specified criteria. Furthermore, the instruction set may be used to associate or link certain functional benefits, such improved characteristics, with one or more identified sequence.

For example, the instruction set can include, e.g., a sequence comparison or other alignment program, e.g., an available program such as, for example, the Wisconsin Package Version 10.0, such as BLAST, FASTA, PILEUP, FINDPATTERNS or the like (GCG, Madison, WI). Public sequence databases such as GenBank, EMBL, Swiss-Prot and PIR or private sequence databases such as PHYTOSEQ sequence database (Incyte Genomics, Palo Alto, CA) can be searched.

Alignment of sequences for comparison can be conducted by the local homology algorithm of Smith and Waterman (1981) *Adv. Appl. Math.* 2: 482-489, by the homology alignment algorithm of Needleman and Wunsch (1970) *J. Mol. Biol.* 48: 443-453, by the search for similarity method of Pearson and Lipman (1988) *Proc. Natl. Acad. Sci.* 85: 2444-2448, by computerized implementations of these algorithms. After alignment, sequence comparisons between two (or more) polynucleotides or polypeptides are typically performed by comparing sequences of the two sequences over a comparison window to identify and compare local regions of sequence similarity. The comparison window can be a segment of at least about 20 contiguous positions, usually about 50 to about 200, more usually about 100 to about 150 contiguous positions. A description of the method is provided in Ausubel et al. *supra*.

A variety of methods for determining sequence relationships can be used, including manual alignment and computer assisted sequence alignment and analysis. This later approach is a preferred approach in the present invention, due to the increased throughput afforded by computer assisted methods. As noted above, a variety of computer programs for performing sequence alignment are available, or can be produced by one of skill.

One example algorithm that is suitable for determining percent sequence identity and sequence similarity is the BLAST algorithm, which is described in Altschul et al. (1990) *J. Mol. Biol.* 215: 403-410. Software for performing BLAST analyses is publicly available, e.g., through the National Library of Medicine's National Center for Biotechnology Information (ncbi.nlm.nih; see at world wide web (www) National Institutes of Health US government (gov) website). This algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence, which either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as the neighborhood word score threshold (Altschul et al. *supra*). These initial neighborhood word hits act as seeds for initiating searches to find longer HSPs containing them. The word hits are then extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Cumulative scores are calculated using, for nucleotide

sequences, the parameters M (reward score for a pair of matching residues; always > 0) and N (penalty score for mismatching residues; always < 0). For amino acid sequences, a scoring matrix is used to calculate the cumulative score. Extension of the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters W, T, and X determine the sensitivity and speed of the alignment. The BLASTN program (for nucleotide sequences) uses as defaults a wordlength (W) of 11, an expectation (E) of 10, a cutoff of 100, M=5, N=-4, and a comparison of both strands. For amino acid sequences, the BLASTP program uses as defaults a wordlength (W) of 3, an expectation (E) of 10, and the BLOSUM62 scoring matrix (*see* Henikoff and Henikoff (1992) *Proc. Natl. Acad. Sci.* 89: 10915-10919). Unless otherwise indicated, "sequence identity" here refers to the % sequence identity generated from a tblastx using the NCBI version of the algorithm at the default settings using gapped alignments with the filter "off" (see, for example, NIH NLM NCBI website at ncbi.nlm.nih, supra).

In addition to calculating percent sequence identity, the BLAST algorithm also performs a statistical analysis of the similarity between two sequences (*see,* e.g. Karlin and Altschul (1993) *Proc. Natl. Acad. Sci.* 90: 5873-5787). One measure of similarity provided by the BLAST algorithm is the smallest sum probability (P(N)), which provides an indication of the probability by which a match between two nucleotide or amino acid sequences would occur by chance. For example, a nucleic acid is considered similar to a reference sequence (and, therefore, in this context, homologous) if the smallest sum probability in a comparison of the test nucleic acid to the reference nucleic acid is less than about 0.1, or less than about 0.01, and or even less than about 0.001. An additional example of a useful sequence alignment algorithm is PILEUP. PILEUP creates a multiple sequence alignment from a group of related sequences using progressive, pairwise alignments. The program can align, e.g., up to 300 sequences of a maximum length of 5,000 letters.

The integrated system, or computer typically includes a user input interface allowing a user to selectively view one or more sequence records corresponding to the one or more character strings, as well as an instruction set which aligns the one or more character strings with each other or with an additional character string to identify one or more region of sequence similarity. The system may include a link of one or more character strings with a particular phenotype or gene function. Typically, the system includes a user readable output element that displays an alignment produced by the alignment instruction set.

The methods of this invention can be implemented in a localized or distributed computing environment. In a distributed environment, the methods may implemented on a single computer comprising multiple processors or on a multiplicity of computers. The computers can be linked, e.g. through a common bus, but more preferably the computer(s) are nodes on a network. The network can be a generalized or a dedicated local or wide-area network and, in certain preferred embodiments, the computers may be components of an intra-net or an internet.

Thus, the invention provides methods for identifying a sequence similar or homologous to one or more polynucleotides as noted herein, or one or more target polypeptides encoded by the polynucleotides, or otherwise noted herein and may include linking or associating a given plant phenotype or gene function with a sequence. In the methods, a sequence database is provided (locally or across an inter or intra net) and a query is made against the sequence database using the relevant sequences herein and associated plant phenotypes or gene functions.

Any sequence herein can be entered into the database, before or after querying the database. This provides for both expansion of the database and, if done before the querying step, for insertion of control sequences into the database. The control sequences can be detected by the query to ensure the general integrity of both the database and the query. As noted, the query can be performed using a web browser based interface. For example, the database can be a centralized public database such as those noted herein, and the querying can be done from a remote terminal or computer across an internet or intranet.

Any sequence herein can be used to identify a similar, homologous, paralogous, or orthologous sequence in another plant. This provides means for identifying endogenous sequences in other plants that may be useful to alter a trait of progeny plants, which results from crossing two plants of different strain. For example, sequences that encode an ortholog of any of the sequences herein that naturally occur in a plant with a desired trait can be identified using the sequences disclosed herein. The plant is then crossed with a second plant of the same species but which does not have the desired trait to produce progeny which can then be used in further crossing experiments to produce the desired trait in the second plant. Therefore the resulting progeny plant contains no transgenes; expression of the endogenous sequence may also be regulated by treatment with a particular chemical or other means, such as EMR. Some examples of such compounds well known in the art include: ethylene; cytokinins; phenolic compounds, which stimulate the transcription of the genes needed for infection; specific monosaccharides and acidic environments which potentiate vir gene induction; acidic polysaccharides which induce one or more chromosomal genes; and opines; other mechanisms include light or dark treatment (for a review of examples of such treatments, see, Winans (1992) *Microbiol. Rev.* 56: 12-31; Eyal et al. (1992) *Plant Mol. Biol.* 19: 589-599; Chrispeels et al. (2000) *Plant Mol. Biol.* 42: 279-290; Piazza et al. (2002) *Plant Physiol.* 128: 1077-1086).

Table 10 lists sequences discovered to be orthologous to a number of representative transcription factors of the present invention. The column headings include the transcription factors listed by SEQ ID NO; corresponding Gene ID (GID) numbers; the species from which the orthologs to the transcription factors are derived; the type of sequence (i.e., DNA or protein) discovered to be orthologous to the transcription factors; and the SEQ ID NO of the orthologs, the latter corresponding to the ortholog SEQ ID NOs listed in the Sequence Listing.

Table 10. Orthologs of Representative *Arabidopsis* Transcription Factor Genes

SEQ ID NO: of Ortholog or Nucleotide Encoding Ortholog	Ortholog GID NO	Species from Which Ortholog is Derived	Sequence type used for determination (DNA or Protein)	GID NO of Orthologous *Arabidopsis* Transcription Factor	SEQ ID NO: of Nucleotide Encoding Orthologous *Arabidopsis* Transcription Factor
468		*Glycine max*	DNA	G19	3
469		*Glycine max*	DNA	G19	3
470		*Glycine max*	DNA	G19	3
471		*Glycine max*	DNA	G19	3
472		*Oryza sativa*	DNA	G19	3
473		*Oryza sativa*	DNA	G19	3
474		*Oryza sativa*	DNA	G19	3
475		*Zea mays*	DNA	G19	3
476		*Zea mays*	DNA	G19	3
477		*Glycine max*	DNA	G22	5
478		*Glycine max*	DNA	G22	5
491		*Glycine max*	DNA	G28	9
492		*Glycine max*	DNA	G28	9
493		*Glycine max*	DNA	G28	9
494		*Glycine max*	DNA	G28	9
495		*Glycine max*	DNA	G28	9
496		*Glycine max*	DNA	G28	9
497		*Glycine max*	DNA	G28	9
498		*Glycine max*	DNA	G28	9
499		*Oryza sativa*	DNA	G28	9
500		*Zea mays*	DNA	G28	9
501		*Oryza sativa*	PRT	G28	9
502		*Oryza sativa*	PRT	G28	9
503		*Mesembryanthemum crystallinum*	PRT	G28	9
504	G3643	*Glycine max*	DNA	G47, G2133	11, 407
505		*Oryza sativa*	PRT	G47, G2133	11, 407
550	G3450	*Glycine max*	DNA	G226, G682	37, 147
551		*Glycine max*	DNA	G226, G682	37
552		*Glycine max*	DNA	G226, G682	37, 147
553	G3448	*Glycine max*	DNA	G226, G682	37, 147
554	G3449	*Glycine max*	DNA	G226, G682	37, 147
555		*Oryza sativa*	DNA	G226, G682	37, 147
556	G3431	*Zea mays*	DNA	G226, G682	37, 147
557		*Zea mays*	DNA	G226, G682	37, 147
558		*Oryza sativa*	PRT	G226, G682	37, 147
559	G3393	*Oryza sativa*	PRT	G226, G682	37, 147
610		*Glycine max*	DNA	G353, G354	59, 61
611		*Glycine max*	DNA	G353, G354	59, 61
612		*Glycine max*	DNA	G353, G354	59, 61
613		*Oryza sativa*	DNA	G353, G354	59, 61
614		*Zea mays*	DNA	G353, G354	59, 61
615		*Zea mays*	DNA	G353, G354	59, 61

616		*Zea mays*	DNA	G353, G354	59, 61
617		*Zea mays*	DNA	G353, G354	59, 61
618		*Zea mays*	DNA	G353, G354	59, 61
619		*Zea mays*	DNA	G353, G354	59, 61
620		*Zea mays*	DNA	G353, G354	59, 61
621		*Oryza sativa*	PRT	G353, G354	59, 61
622		*Oryza sativa*	PRT	G353, G354	59, 61
623		*Oryza sativa*	PRT	G353, G354	59, 61
624		*Oryza sativa*	PRT	G353, G354	59, 61
625		*Oryza sativa*	PRT	G353, G354	59, 61
626		*Oryza sativa*	PRT	G353, G354	59, 61
746		*Glycine max*	DNA	G481, G482	87, 89
747		*Glycine max*	DNA	G481, G482	87, 89
748		*Glycine max*	DNA	G481, G482	87, 89
749	G3476	*Glycine max*	DNA	G481, G482	87, 89
750		*Glycine max*	DNA	G481, G482	87, 89
751		*Glycine max*	DNA	G481, G482	87, 89
752		*Glycine max*	DNA	G481, G482	87, 89
753		*Glycine max*	DNA	G481, G482	87, 89
754		*Glycine max*	DNA	G481, G482	87
755		*Glycine max*	DNA	G481, G482	87
756		*Oryza sativa*	DNA	G481, G482	87
757		*Oryza sativa*	DNA	G481, G482	87, 89
758		*Zea mays*	DNA	G481, G482	87
759		*Zea mays*	DNA	G481, G482	87, 89
760		*Zea mays*	DNA	G481, G482	87, 89
761		*Zea mays*	DNA	G481, G482	87, 89
762		*Zea mays*	DNA	G481, G482	87, 89
763		*Zea mays*	DNA	G481, G482	87, 89
764		*Zea mays*	DNA	G481, G482	87, 89
765		*Zea mays*	DNA	G481, G482	87, 89
766		*Zea mays*	DNA	G481, G482	87, 89
767		*Zea mays*	DNA	G481, G482	87, 89
768		*Gossypium arboreum*	DNA	G481, G482	87, 89
769		*Glycine max*	DNA	G481, G482	87, 89
770		*Gossypium hirsutum*	DNA	G481, G482	87, 89
771		*Lycopersicon esculentum*	DNA	G481, G482	87, 89
772		*Lycopersicon esculentum*	DNA	G481, G482	87, 89
773		*Medicago truncatula*	DNA	G481, G482	87, 89
774		*Lycopersicon esculentum*	DNA	G481, G482	87, 89
775		*Solanum tuberosum*	DNA	G481, G482	87, 89
776		*Triticum aestivum*	DNA	G481, G482	87, 89
777		*Hordeum vulgare*	DNA	G481, G482	87, 89
778		*Triticum monococcum*	DNA	G481, G482	87, 89
779		*Glycine max*	DNA	G481, G482	89

780		*Oryza sativa*	PRT	G481, G482	87, 89
781		*Oryza sativa*	PRT	G481, G482	87, 89
782		*Oryza sativa*	PRT	G481, G482	87, 89
783		*Oryza sativa*	PRT	G481, G482	87, 89
784		*Oryza sativa*	PRT	G481, G482	87, 89
785		*Zea mays*	PRT	G481, G482	87, 89
786		*Zea mays*	PRT	G481, G482	87, 89
787		*Oryza sativa*	PRT	G481, G482	87, 89
788		*Oryza sativa*	PRT	G481, G482	87, 89
789		*Oryza sativa*	PRT	G481, G482	87, 89
790	G3395	*Oryza sativa*	PRT	G481, G482	87, 89
791		*Oryza sativa*	PRT	G481, G482	87, 89
792		*Oryza sativa*	PRT	G481, G482	87, 89
793		*Oryza sativa*	PRT	G481, G482	87, 89
794	G3397	*Oryza sativa*	PRT	G481, G482	87, 89
795		*Oryza sativa*	PRT	G481, G482	87, 89
796	G3398	*Oryza sativa*	PRT	G481, G482	87, 89
797		*Glycine max*	PRT	G481, G482	87, 89
798	G3476	*Glycine max*	PRT	G481, G482	87, 89
799		*Glycine max*	PRT	G481, G482	87, 89
800	G3475	*Glycine max*	PRT	G481, G482	87, 89
801	G3472	*Glycine max*	PRT	G481, G482	87, 89
802		*Glycine max*	PRT	G481, G482	87, 89
803		*Glycine max*	PRT	G481, G482	87, 89
804		*Zea mays*	PRT	G481, G482	87, 89
805	G3436	*Zea mays*	PRT	G481, G482	87, 89
806	G3434	*Zea mays*	PRT	G481, G482	87, 89
807		*Zea mays*	PRT	G481, G482	87, 89
825		*Glycine max*	DNA	G489	93
826		*Glycine max*	DNA	G489	93
827		*Glycine max*	DNA	G489	93
828		*Glycine max*	DNA	G489	93
829		*Glycine max*	DNA	G489	93
830		*Glycine max*	DNA	G489	93
831		*Glycine max*	DNA	G489	93
832		*Oryza sativa*	DNA	G489	93
833		*Oryza sativa*	DNA	G489	93
834		*Zea mays*	DNA	G489	93
835		*Oryza sativa*	PRT	G489	93
836		*Oryza sativa*	PRT	G489	93
837		*Oryza sativa*	PRT	G489	93
981		*Oryza sativa*	DNA	G634	127
982		*Oryza sativa*	DNA	G634	127
983		*Oryza sativa*	DNA	G634	127
984		*Zea mays*	DNA	G634	127
985		*Zea mays*	DNA	G634	127
986		*Zea mays*	DNA	G634	127
987		*Oryza sativa*	PRT	G634	127
988		*Oryza sativa*	PRT	G634	127
1076	G3450	*Glycine max*	DNA	G682	147

1077		*Hordeum vulgare subsp. vulgare*	DNA	G682	147
1078		*Populus tremula x Populus tremuloides*	DNA	G682	147
1079		*Triticum aestivum*	DNA	G682	147
1080		*Gossypium arboreum*	DNA	G682	147
1081		*Oryza sativa*	PRT	G682	147
1082	G3392	*Oryza sativa*	PRT	G682	147
1083	G3445	*Glycine max*	PRT	G682	147
1084	G3450	*Glycine max*	PRT	G682	147
1085		*Glycine max*	PRT	G682	147
1086	G3446	*Glycine max*	PRT	G682	147
1087	G3448	*Glycine max*	PRT	G682	147
1088	G3449	*Glycine max*	PRT	G682	147
1089	G3431	*Zea mays*	PRT	G682	147
1090		*Zea mays*	PRT	G682	147
1154		*Glycine max*	DNA	G864	167
1155		*Glycine max*	DNA	G864	167
1156		*Zea mays*	DNA	G864	167
1157		*Oryza sativa*	PRT	G864	167
1158		*Oryza sativa*	PRT	G864	167
1159		*Glycine max*	DNA	G867, G1930	169, 369
1160	G3451	*Glycine max*	DNA	G867, G1930	169, 369
1161		*Glycine max*	DNA	G867, G1930	169, 369
1162	G3452	*Glycine max*	DNA	G867, G1930	169, 369
1163		*Glycine max*	DNA	G867, G1930	169, 369
1164		*Glycine max*	DNA	G867, G1930	169
1165		*Oryza sativa*	DNA	G867, G1930	169
1166		*Oryza sativa*	DNA	G867, G1930	169, 369
1167		*Zea mays*	DNA	G867, G1930	169, 369
1168		*Zea mays*	DNA	G867, G1930	169, 369
1169		*Zea mays*	DNA	G867, G1930	169, 369
1170		*Zea mays*	DNA	G867, G1930	169, 369
1171		*Glycine max*	DNA	G867, G1930	169, 369
1172		*Mesembryanthemum crystallinum*	DNA	G867, G1930	169, 369
1173		*Lycopersicon esculentum*	DNA	G867, G1930	169, 369
1174		*Solanum tuberosum*	DNA	G867, G1930	169, 369
1175		*Hordeum vulgare*	DNA	G867, G1930	169, 369
1176	G3388	*Oryza sativa*	PRT	G867, G1930	169, 369
1177	G3389	*Oryza sativa*	PRT	G867, G1930	169, 369
1178		*Oryza sativa*	PRT	G867, G1930	169, 369
1179	G3390	*Oryza sativa*	PRT	G867, G1930	169, 369
1180		*Oryza sativa*	PRT	G867, G1930	169, 369
1181		*Oryza sativa*	PRT	G867, G1930	169, 369
1182	G3451	*Glycine max*	PRT	G867, G1930	169, 369
1183	G3452	*Glycine max*	PRT	G867, G1930	169, 369
1184	G3453	*Glycine max*	PRT	G867, G1930	169, 369
1185	G3433	*Zea mays*	PRT	G867, G1930	169, 369

1186		*Zea mays*	PRT	G867, G1930	169, 369
1204		*Glycine max*	DNA	G912	185
1205		*Glycine max*	DNA	G912	185
1206		*Glycine max*	DNA	G912	185
1207		*Glycine max*	DNA	G912	185
1208		*Glycine max*	DNA	G912	185
1209		*Glycine max*	DNA	G912	185
1210		*Glycine max*	DNA	G912	185
1211		*Oryza sativa*	DNA	G912	185
1212		*Oryza sativa*	DNA	G912, G913	185, 187
1213	G3440	*Zea mays*	DNA	G912	185
1214		*Zea mays*	DNA	G912	185
1215		*Zea mays*	DNA	G912, G913	185, 187
1216		*Zea mays*	DNA	G912	185
1217		*Zea mays*	DNA	G912	185
1218		*Brassica napus*	DNA	G912, G913	185, 187
1219		*Solanum tuberosum*	DNA	G912	185
1220		*Descurainia sophia*	DNA	G912	185
1221	G3377	*Oryza sativa*	PRT	G912	185
1222	G3373	*Oryza sativa*	PRT	G912, G913	185, 187
1223	G3375	*Oryza sativa*	PRT	G912, G913	185, 187
1224	G3372	*Oryza sativa*	PRT	G912	185
1225		*Brassica napus*	PRT	G912	185
1226		*Nicotiana tabacum*	PRT	G912	185
1227		*Oryza sativa*	PRT	G912	185
1228		*Oryza sativa*	PRT	G912	185
1229		*Oryza sativa*	PRT	G912	185
1230	G3379	*Oryza sativa*	PRT	G912	185
1231		*Oryza sativa*	PRT	G912	185
1232	G3376	*Oryza sativa*	PRT	G912	185
1233		*Oryza sativa*	PRT	G912	185
1234		*Oryza sativa*	PRT	G912	185
1235		*Oryza sativa*	PRT	G912	185
1236		*Oryza sativa*	PRT	G912	185
1237		*Glycine max*	PRT	G912	185
1238		*Glycine max*	PRT	G912	185
1239		*Glycine max*	PRT	G912	185
1240		*Glycine max*	PRT	G912	185
1241		*Glycine max*	PRT	G912	185
1242		*Glycine max*	PRT	G912	185
1243		*Glycine max*	PRT	G912	185
1244		*Zea mays*	PRT	G912	185
1245		*Zea mays*	PRT	G912	185
1246	G3440	*Zea mays*	PRT	G912	185
1247		*Zea mays*	PRT	G912	185
1248		*Zea mays*	PRT	G912	185
1249		*Glycine max*	DNA	G922	189
1250	G3811	*Glycine max*	DNA	G922	189
1251		*Glycine max*	DNA	G922	189
1252		*Oryza sativa*	DNA	G922	189

1253		*Oryza sativa*	DNA	G922	189
1254		*Oryza sativa*	PRT	G922	189
1255		*Oryza sativa*	PRT	G922	189
1256		*Oryza sativa*	PRT	G922	189
1257	G3813	*Oryza sativa*	PRT	G922	189
1258		*Glycine max*	DNA	G926	191
1259		*Glycine max*	DNA	G926	191
1260		*Oryza sativa*	DNA	G926	191
1261		*Oryza sativa*	DNA	G926	191
1262		*Zea mays*	DNA	G926	191
1263		*Brassica napus*	PRT	G926	191
1292		*Glycine max*	DNA	G975, G2583	199, 449
1293		*Glycine max*	DNA	G975, G2583	199, 449
1294		*Glycine max*	DNA	G975, G2583	199, 449
1295		*Glycine max*	DNA	G975, G2583	199, 449
1296		*Glycine max*	DNA	G975, G2583	199, 449
1297		*Oryza sativa*	DNA	G975	199
1298		*Oryza sativa*	DNA	G975, G2583	199, 449
1299		*Zea mays*	DNA	G975, G2583	199, 449
1300		*Zea mays*	DNA	G975, G2583	199, 449
1301		*Brassica rapa*	DNA	G975, G2583	199, 449
1302		*Oryza sativa*	PRT	G975, G2583	199, 449
1393		*Glycine max*	DNA	G1069, G2153	221, 417
1394		*Glycine max*	DNA	G1069, G2153	221, 417
1395		*Oryza sativa*	PRT	G1069, G1073, G2153	221, 223, 417
1396		*Zea mays*	DNA	G1069, G2153	221, 417
1397		*Lotus japonicus*	DNA	G1069, G2153	221, 417
1398		*Lycopersicon esculentum*	DNA	G1073	223
1399	G3399	*Oryza sativa*	PRT	G1073	223
1400		*Oryza sativa*	PRT	G1073	223
1401	G3400	*Oryza sativa*	PRT	G1073	223
1402		*Oryza sativa*	PRT	G1073	223
1403		*Oryza sativa*	PRT	G1073	223
1404		*Oryza sativa*	PRT	G1073	223
1405		*Oryza sativa*	PRT	G1073	223
1406		*Oryza sativa*	PRT	G1073	223
1407		*Oryza sativa*	PRT	G1073	223
1408		*Oryza sativa*	PRT	G1073	223
1409		*Oryza sativa*	PRT	G1073	223
1410		*Oryza sativa*	PRT	G1073	223
1411		*Glycine max*	PRT	G1073	223
1412		*Glycine max*	PRT	G1073	223
1413		*Glycine max*	PRT	G1073	223
1414		*Glycine max*	PRT	G1073	223
1415		*Glycine max*	PRT	G1073	223
1416		*Glycine max*	PRT	G1073	223
1417		*Glycine max*	PRT	G1073	223
1418		*Zea mays*	PRT	G1073	223
1419		*Glycine max*	DNA	G1075	225

1420		*Glycine max*	DNA	G1075	225
1421		*Glycine max*	DNA	G1075	225
1422		*Glycine max*	DNA	G1075	225
1423		*Glycine max*	DNA	G1075	225
1424		*Oryza sativa*	DNA	G1075	225
1425		*Oryza sativa*	DNA	G1075	225
1426		*Oryza sativa*	DNA	G1075	225
1587		*Glycine max*	DNA	G1411, G2509	269, 439
1588		*Glycine max*	DNA	G1411, G2509	269, 439
1589		*Glycine max*	DNA	G1411, G2509	269, 439
1590		*Glycine max*	DNA	G1411, G2509	269, 439
1591		*Zea mays*	DNA	G1411, G2509	269, 439
1604		*Glycine max*	DNA	G1451	277
1605		*Glycine max*	DNA	G1451	277
1606		*Oryza sativa*	DNA	G1451	277
1607		*Oryza sativa*	DNA	G1451	277
1608		*Oryza sativa*	DNA	G1451	277
1609		*Zea mays*	DNA	G1451	277
1610		*Zea mays*	DNA	G1451	277
1611		*Zea mays*	DNA	G1451	277
1612		*Zea mays*	DNA	G1451	277
1613		*Medicago truncatula*	DNA	G1451	277
1614		*Solanum tuberosum*	DNA	G1451	277
1615		*Zea mays*	DNA	G1451	277
1616		*Sorghum propinquum*	DNA	G1451	277
1617		*Glycine max*	DNA	G1451	277
1618		*Sorghum bicolor*	DNA	G1451	277
1619		*Hordeum vulgare*	DNA	G1451	277
1620		*Lycopersicon esculentum*	DNA	G1451	277
1621		*Oryza sativa*	PRT	G1451	277
1622		*Oryza sativa*	PRT	G1451	277
1623		*Oryza sativa*	PRT	G1451	277
1624		*Oryza sativa*	PRT	G1451	277
1671		*Glycine max*	DNA	G1543	303
1672		*Oryza sativa*	DNA	G1543	303
1673		*Zea mays*	DNA	G1543	303
1674		*Oryza sativa*	PRT	G1543	303
1728		*Glycine max*	DNA	G1792	331
1729		*Glycine max*	DNA	G1792	331
1730		*Glycine max*	DNA	G1792	331
1731		*Glycine max*	DNA	G1792	331
1732		*Glycine max*	DNA	G1792	331
1733		*Zea mays*	DNA	G1792	331
1734		*Lycopersicon esculentum*	DNA	G1792	331
1735	G3380	*Oryza sativa*	PRT	G1792	331
1736	G3381	*Oryza sativa indica*	PRT	G1792	331
1737	G3383	*Oryza sativa*	PRT	G1792	331

		japonica			
1795		*Oryza sativa*	DNA	G1930	369
1908		*Medicago truncatula*	DNA	G2155	419
1909		*Medicago truncatula*	DNA	G2155	419
1910		*Glycine max*	DNA	G2155	419
2907	G3472	*Glycine max*	DNA	G481, G482	87, 89
2908	G3475	*Glycine max*	DNA	G481, G482	87, 89
2909	G3476	*Glycine max*	DNA	G481, G482	87, 89
2910	G3395	*Oryza sativa*	DNA	G481, G482	87, 89
2911	G3397	*Oryza sativa*	DNA	G481, G482	87, 89
2912	G3398	*Oryza sativa*	DNA	G481, G482	87, 89
2913	G3434	*Zea mays*	DNA	G481, G482	87, 89
2914	G3436	*Zea mays*	DNA	G481, G482	87, 89
2915	G3445	*Glycine max*	DNA	G682	147
2916	G3446	*Glycine max*	DNA	G682	147
2917	G3448	*Glycine max*	DNA	G682	147
2918	G3449	*Glycine max*	DNA	G682	147
2919	G3450	*Glycine max*	DNA	G682	147
2920	G3392	*Oryza sativa*	DNA	G682	147
2921	G3393	*Oryza sativa*	DNA	G226, G682	37, 147
2922	G3431	*Zea mays*	DNA	G682	147
2923	G3451	*Glycine max*	DNA	G867, G1930	169, 369
2924	G3452	*Glycine max*	DNA	G867, G1930	169, 369
2925	G3453	*Glycine max*	DNA	G867, G1930	169, 369
2926	G3388	*Oryza sativa*	DNA	G867, G1930	169, 369
2927	G3389	*Oryza sativa*	DNA	G867, G1930	169, 369
2928	G3390	*Oryza sativa*	DNA	G867, G1930	169, 369
2929	G3433	*Zea mays*	DNA	G867, G1930	169, 369
2930	G3376	*Oryza sativa*	DNA	G912	185
2931	G3372	*Oryza sativa*	DNA	G912	185
2932	G3373	*Oryza sativa*	DNA	G912, G913	185, 187
2933	G3375	*Oryza sativa*	DNA	G912, G913	185, 187
2934	G3377	*Oryza sativa*	DNA	G912	185
2935	G3379	*Oryza sativa*	DNA	G912	185
2936	G3440	*Zea mays*	DNA	G912	185
2937	G3399	*Oryza sativa*	DNA	G1073	223
2938	G3400	*Oryza sativa*	DNA	G1073	223
2939	G3380	*Oryza sativa*	DNA	G1792	331
2940	G3381	*Oryza sativa indica*	DNA	G1792	331
2941	G3383	*Oryza sativa japonica*	DNA	G1792	331
2942	G3643	*Glycine max*	PRT	G47, G2133	11, 407
2943	G3811	*Glycine max*	PRT	G922	189
2944	G3813	*Oryza sativa*	DNA	G922	189
2945	G3429	*Oryza sativa*	DNA	G481, G482	87, 89
2946	G3429	*Oryza sativa*	PRT	G481, G482	87, 89
2947	G3470	*Glycine max*	DNA	G481, G482	87, 89
2948	G3470	*Glycine max*	PRT	G481, G482	87, 89
2949	G3471	*Glycine max*	DNA	G481, G482	87, 89

2950	G3471	*Glycine max*	PRT	G481, G482	87, 89

Table 11 lists a summary of homologous sequences identified using BLAST (tblastx program). The first column shows the polynucleotide sequence identifier (SEQ ID NO), the second column shows the corresponding cDNA identifier (Gene ID), the third column shows the orthologous or homologous polynucleotide GenBank Accession Number (Test Sequence ID), the fourth column shows the calculated probability value that the sequence identity is due to chance (Smallest Sum Probability), the fifth column shows the plant species from which the test sequence was isolated (Test Sequence Species), and the sixth column shows the orthologous or homologous test sequence GenBank annotation (Test Sequence GenBank Annotation).

Table 11. Summary of representative sequences that are homologous to presently disclosed transcription factors

SEQ ID NO:	GID	Test Sequence ID	Smallest Sum Probability	Test Sequence Species	Test Sequence GenBank Annotation
3	G19	BG321358	1.00E-101	Descurainia sophia	Ds01_07d03_R Ds01_AAFC_ECORC_cold_ stress
3	G19	BH444831	1.00E-77	Brassica oleracea	BOHPW42TR BOHP Brassica oleracea genomic
3	G19	BM412184	2.00E-43	Lycopersicon esculentum	EST586511 tomato breaker fruit Lyco
3	G19	BU837697	3.00E-43	Populus tremula x Populus tremuloides	T104G02 Populus apica
3	G19	CA784650	6.00E-43	Glycine max	sat87a10.y1 Gm-c1062 Glycine max cDNA clone SOY
3	G19	BU819833	3.00E-41	Populus tremula	UA48BPB07 Populus tremula cambium cDNA libr
3	G19	BU870388	4.00E-41	Populus balsamifera subsp. trichocarpa	Q011H05 Populus flow
3	G19	CA797119	1.00E-38	Theobroma cacao	Cac_BL_4204 Cac_BL (Bean and Leaf from Amel
3	G19	BI436183	2.00E-38	Solanum tuberosum	EST538944 cSTE Solanum tuberosum cDNA clo
3	G19	BQ989448	2.00E-36	Lactuca sativa	QGF17L05.yg.ab1 QG_EFGHJ lettuce serriola La
3	G19	gi10798644	5.70E-36	Nicotiana tabacum	AP2 domain-containing transcription fac
3	G19	gi6176534	2.40E-35	Oryza sativa	EREBP-like protein.
3	G19	gi1688233	7.50E-34	Solanum tuberosum	DNA binding protein homolog.
3	G19	gi22074046	1.50E-33	Lycopersicon esculentum	transcription factor JERF1.

3	G19	gi18496063	4.90E-33	Fagus sylvatica	ethylene responsive element binding prote
3	G19	gi20805105	2.10E-32	Oryza sativa (japonica cultivar-group)	contains ESTs AU06
3	G19	gi24940524	2.30E-31	Triticum aestivum	ethylene response element binding prote
3	G19	gi18266198	2.30E-31	Narcissus pseudonarcissus	AP-2 domain containing protein.
3	G19	gi3264767	1.30E-30	Prunus armeniaca	AP2 domain containing protein.
3	G19	gi24817250	4.00E-28	Cicer arietinum	transcription factor EREBP-like protein.
5	G22	AB016264	9.00E-48	Nicotiana sylvestris	nserf2 gene for ethylene-responsive el
5	G22	TOBBY4A	1.00E-47	Nicotiana tabacum	mRNA for ERF1, complete cds.
5	G22	AP004533	4.00E-47	Lotus japonicus	genomic DNA, chromosome 3, clone:LjT14G02,
5	G22	LEU89255	6.00E-47	Lycopersicon esculentum	DNA-binding protein Pti4 mRNA, comp
5	G22	BQ517082	6.00E-46	Solanum tuberosum	EST624497 Generation of a set of potato c
5	G22	BE449392	1.00E-45	Lycopersicon hirsutum	EST356151 L. hirsutum trichome, Corne
5	G22	AF245119	5.00E-45	Mesembryanthemum crystallinum	AP2-related transcription fac
5	G22	BQ165291	7.00E-45	Medicago truncatula	EST611160 KVKC Medicago truncatula cDNA
5	G22	AW618245	8.00E-38	Lycopersicon pennellii	EST314295 L. pennellii trichome, Cor
5	G22	BG444654	2.00E-36	Gossypium arboreum	GA__Ea0025B11f Gossypium arboreum 7-10 d
5	G22	gi1208495	6.10E-48	Nicotiana tabacum	ERF1.
5	G22	gi3342211	3.30E-47	Lycopersicon esculentum	Pti4.
5	G22	gi8809571	8.90E-47	Nicotiana sylvestris	ethylene-responsive element binding
5	G22	gi17385636	2.70E-36	Matricaria chamomilla	ethylene-responsive element binding
5	G22	gi8980313	2.50E-33	Catharanthus roseus	AP2-domain DNA-binding protein.
5	G22	gi7528276	8.60E-33	Mesembryanthemum crystallinum	AP2-related transcription f
5	G22	gi21304712	3.10E-28	Glycine max	ethylene-responsive element binding protein 1
5	G22	gi14140141	1.50E-26	Oryza sativa	putative AP2-related transcription factor.
5	G22	gi15623863	1.30E-22	Oryza sativa (japonica cultivar-group)	contains EST~hypot
5	G22	gi4099914	3.10E-21	Stylosanthes hamata	ethylene-responsive element binding p

9	G28	AF245119	2.00E-72	Mesembryanthemum crystallinum	AP2-related transcription fac
9	G28	BQ165291	1.00E-68	Medicago truncatula	EST611160 KVKC Medicago truncatula cDNA
9	G28	AB016264	1.00E-57	Nicotiana sylvestris	nserf2 gene for ethylene-responsive el
9	G28	TOBBY4D	2.00E-57	Nicotiana tabacum	Tobacco mRNA for EREBP-2, complete cds.
9	G28	BQ047502	2.00E-57	Solanum tuberosum	EST596620 P. infestans-challenged potato
9	G28	LEU89255	2.00E-56	Lycopersicon esculentum	DNA-binding protein Pti4 mRNA, comp
9	G28	BH454277	2.00E-54	Brassica oleracea	BOGSI45TR BOGS Brassica oleracea genomic
9	G28	BE449392	1.00E-53	Lycopersicon hirsutum	EST356151 L. hirsutum trichome, Corne
9	G28	AB035270	2.00E-50	Matricaria chamomilla	McEREBP1 mRNA for ethylene-responsive
9	G28	AW233956	5.00E-50	Glycine max	sf32e02.y1 Gm-c1028 Glycine max cDNA clone GENO
9	G28	gi7528276	6.10E-71	Mesembryanthemum crystallinum	AP2-related transcription f
9	G28	gi8809571	3.30E-56	Nicotiana sylvestris	ethylene-responsive element binding
9	G28	gi3342211	4.20E-56	Lycopersicon esculentum	Pti4.
9	G28	gi1208498	8.70E-56	Nicotiana tabacum	EREBP-2.
9	G28	gi14140141	4.20E-49	Oryza sativa	putative AP2-related transcription factor.
9	G28	gi17385636	3.00E-46	Matricaria chamomilla	ethylene-responsive element binding
9	G28	gi21304712	2.90E-31	Glycine max	ethylene-responsive element binding protein 1
9	G28	gi15623863	5.60E-29	Oryza sativa (japonica cultivar-group)	contains EST~hypot
9	G28	gi8980313	1.20E-26	Catharanthus roseus	AP2-domain DNA-binding protein.
9	G28	gi4099921	3.10E-21	Stylosanthes hamata	EREBP-3 homolog.
11	G47	BG543936	1.00E-60	Brassica rapa subsp. pekinensis	E1686 Chinese cabbage etiol
11	G47	BH420519	3.00E-43	Brassica oleracea	BOGUH88TF BOGU Brassica oleracea genomic
11	G47	AU292603	3.00E-30	Zinnia elegans	AU292603 zinnia cultured mesophyll cell equa
11	G47	BE320193	1.00E-24	Medicago truncatula	NF024B04RT1F1029 Developing root Medica
11	G47	AAAA01000718	1.00E-22	Oryza sativa (indica cultivar-group)	( ) scaffold000718
11	G47	AP003379	2.00E-22	Oryza sativa	chromosome 1 clone P0408G07, *** SEQUENCING IN

11	G47	AC124836	8.00E-21	Oryza sativa (japonica cultivar-group)	() chromosome 5 clo
11	G47	BZ403609	2.00E-20	Zea mays	OGABN17TM ZM_0.7_1.5_KB Zea mays genomic clone ZMM
11	G47	BM112772	6.00E-17	Solanum tuberosum	EST560308 potato roots Solanum tuberosum
11	G47	BQ698717	1.00E-16	Pinus taeda	NXPV_148_C06_F NXPV (Nsf Xylem Planings wood Ve
11	G47	gi20161239	6.90E-24	Oryza sativa (japonica cultivar-group)	hypothetical prote
11	G47	gi14140155	6.80E-17	Oryza sativa	putative AP2 domain transcription factor.
11	G47	gi21908034	7.00E-15	Zea mays	DRE binding factor 2.
11	G47	gi20303011	1.90E-14	Brassica napus	CBF-like protein CBF5.
11	G47	gi8571476	3.00E-14	Atriplex hortensis	apetala2 domain-containing protein.
11	G47	gi8980313	2.10E-13	Catharanthus roseus	AP2-domain DNA-binding protein.
11	G47	gi19071243	4.40E-13	Hordeum vulgare	CRT/DRE binding factor 1.
11	G47	gi18650662	5.60E-13	Lycopersicon esculentum	ethylene response factor 1.
11	G47	gi17385636	1.20E-12	Matricaria chamomilla	ethylene-responsive element binding
11	G47	gi1208498	1.50E-12	Nicotiana tabacum	EREBP-2.
37	G226	BU872107	2.00E-21	Populus balsamifera subsp. trichocarpa	Q039C07 Populus flow
37	G226	BU831849	2.00E-21	Populus tremula x Populus tremuloides	T026E01 Populus apica
37	G226	BM437313	9.00E-21	Vitis vinifera	VVA017F06_54121 An expressed sequence tag da
37	G226	BI699876	1.00E-19	Glycine max	sag49b09.y1 Gm-c1081 Glycine max cDNA clone GEN
37	G226	AL750151	4.00E-16	Pinus pinaster	AL750151 AS Pinus pinaster cDNA clone AS06C1
37	G226	CA744013	2.00E-12	Triticum aestivum	wri1s.pk006.m22 wri1s Triticum aestivum c
37	G226	BH961028	3.00E-12	Brassica oleracea	odj30d06.g1 B.oleracea002 Brassica olerac
37	G226	BJ472717	8.00E-12	Hordeum vulgare subsp. vulgare	BJ472717 K. Sato unpublished
37	G226	BF617445	8.00E-12	Hordeum vulgare	HVSMEc0017G08f Hordeum vulgare seedling sho
37	G226	CA762299	2.00E-11	Oryza sativa (indica cultivar-group)	BR060003B10F03.ab1 IRR
37	G226	gi9954118	2.20E-11	Solanum tuberosum	tuber-specific and sucrose-responsive e
37	G226	gi14269333	2.50E-10	Gossypium raimondii	myb-like transcription factor

					Myb 3.
37	G226	gi14269335	2.50E-10	Gossypium herbaceum	myb-like transcription factor Myb 3.
37	G226	gi14269337	2.50E-10	Gossypium hirsutum	myb-like transcription factor Myb 3.
37	G226	gi23476297	2.50E-10	Gossypioides kirkii	myb-like transcription factor 3.
37	G226	gi15082210	8.50E-10	Fragaria x ananassa	transcription factor MYB1.
37	G226	gi19072770	8.50E-10	Oryza sativa	typical P-type R2R3 Myb protein.
37	G226	gi15042108	1.40E-09	Zea mays subsp. parviglumis	CI protein.
37	G226	gi15042124	1.40E-09	Zea luxurians	CI protein.
37	G226	gi20514371	1.40E-09	Cucumis sativus	werewolf.
59	G353	BQ790831	5.00E-68	Brassica rapa subsp. pekinensis	E4675 Chinese cabbage etiol
59	G353	BZ019752	1.00E-67	Brassica oleracea	oed85c06.g1 B.oleracea002 Brassica olerac
59	G353	L46574	6.00E-40	Brassica rapa	BNAF1975 Mustard flower buds Brassica rapa cD
59	G353	AB006601	7.00E-26	Petunia x hybrida	mRNA for ZPT2-14, complete cds.
59	G353	BM437146	2.00E-25	Vitis vinifera	VVA015A06_53787 An expressed sequence tag da
59	G353	BI422808	1.00E-24	Lycopersicon esculentum	EST533474 tomato callus, TAMU Lycop
59	G353	BU867080	1.00E-24	Populus tremula x Populus tremuloides	S074B01 Populus imbib
59	G353	BM527789	3.00E-23	Glycine max	sal65h07.y1 Gm-c1061 Glycine max cDNA clone SOY
59	G353	BQ980246	5.00E-23	Lactuca sativa	QGE10I12.yg.ab1 QG_EFGHJ lettuce serriola La
59	G353	BQ121106	2.00E-22	Solanum tuberosum	EST606682 mixed potato tissues Solanum tu
59	G353	gi2346976	6.50E-28	Petunia x hybrida	ZPT2-13.
59	G353	gi15623820	4.40E-25	Oryza sativa	hypothetical protein.
59	G353	gi21104613	1.40E-18	Oryza sativa (japonica cultivar-group)	contains ESTs AU07
59	G353	gi485814	3.10E-13	Triticum aestivum	WZF1.
59	G353	gi7228329	4.00E-12	Medicago sativa	putative TFIIIA (or kruppel)-like zinc fi
59	G353	gi1763063	1.70E-11	Glycine max	SCOF-1.
59	G353	gi2981169	2.60E-11	Nicotiana tabacum	osmotic stress-induced zinc-finger prot
59	G353	gi4666360	1.10E-10	Datisca glomerata	zinc-finger protein 1.
59	G353	gi2129892	2.30E-08	Pisum sativum	probable finger protein Pszf1 - garden pea.
59	G353	gi2058504	0.00018	Brassica rapa	zinc-finger protein-1.
61	G354	BZ083260	5.00E-49	Brassica oleracea	lle29f02.g1 B.oleracea002 Brassica olerac

61	G354	BQ790831	8.00E-45	Brassica rapa subsp. pekinensis	E4675 Chinese cabbage etiol
61	G354	AB006600	6.00E-27	Petunia x hybrida	mRNA for ZPT2-13, complete cds.
61	G354	L46574	1.00E-26	Brassica rapa	BNAF1975 Mustard flower buds Brassica rapa cD
61	G354	BM437146	3.00E-24	Vitis vinifera	VVA015A06_53787 An expressed sequence tag da
61	G354	BQ121105	6.00E-24	Solanum tuberosum	EST606681 mixed potato tissues Solanum tu
61	G354	BM527789	2.00E-23	Glycine max	sal65h07.y1 Gm-c1061 Glycine max cDNA clone SOY
61	G354	AI898309	2.00E-23	Lycopersicon esculentum	EST267752 tomato ovary, TAMU Lycope
61	G354	BU867080	5.00E-22	Populus tremula x Populus tremuloides	S074B01 Populus imbib
61	G354	BQ980246	1.00E-21	Lactuca sativa	QGE10I12.yg.ab1 QG_EFGHJ lettuce serriola La
61	G354	gi2346976	5.60E-29	Petunia x hybrida	ZPT2-13.
61	G354	gi15623820	1.90E-22	Oryza sativa	hypothetical protein.
61	G354	gi21104613	4.00E-19	Oryza sativa (japonica cultivar-group)	contains ESTs AU07
61	G354	gi2981169	1.80E-17	Nicotiana tabacum	osmotic stress-induced zinc-finger prot
61	G354	gi1763063	4.10E-16	Glycine max	SCOF-1.
61	G354	gi4666360	8.90E-15	Datisca glomerata	zinc-finger protein 1.
61	G354	gi2058504	1.00E-14	Brassica rapa	zinc-finger protein-1.
61	G354	gi7228329	4.90E-14	Medicago sativa	putative TFIIIA (or kruppel)-like zinc fi
61	G354	gi485814	3.20E-13	Triticum aestivum	WZF1.
61	G354	gi2129892	1.20E-06	Pisum sativum	probable finger protein Pszf1 - garden pea.
87	G481	BU238020	9.00E-71	Descurainia sophia	Ds01_14a12_A Ds01_AAFC_ECORC_cold_ stress
87	G481	BG440251	2.00E-56	Gossypium arboreum	GA__Ea0006K20f Gossypium arboreum 7-10 d
87	G481	BF071234	1.00E-54	Glycine max	st06h05.y1 Gm-c1065 Glycine max cDNA clone GENO
87	G481	BQ799965	2.00E-54	Vitis vinifera	EST 2134 Green Grape berries Lambda Zap II L
87	G481	BQ488908	5.00E-53	Beta vulgaris	95-E9134-006-006-M23-T3 Sugar beet MPIZ-ADIS-
87	G481	BU499457	1.00E-52	Zea mays	946175D02.y1 946 - tassel primordium prepared by S
87	G481	AI728916	2.00E-52	Gossypium hirsutum	BNLGHi12022 Six-day Cotton fiber Gossypi
87	G481	BG642751	3.00E-52	Lycopersicon esculentum	EST510945 tomato shoot/meristem Lyc

87	G481	BQ857127	3.00E-51	Lactuca sativa	QGB6K24.yg.ab1 QG_ABCDI lettuce salinas Lact
87	G481	BE413647	6.00E-51	Triticum aestivum	SCU001.E10.R990714 ITEC SCU Wheat Endospe
87	G481	gi115840	1.90E-51	Zea mays	CCAAT-BINDING TRANSCRIPTION FACTOR SUBUNIT A (CB
87	G481	gi20160792	2.60E-47	Oryza sativa (japonica cultivar-group)	putative CAAT-box
87	G481	gi15408794	7.10E-38	Oryza sativa	putative CCAAT-binding transcription factor
89	G482	BQ505706	7.00E-59	Solanum tuberosum	EST613121 Generation of a set of potato c
89	G482	AC122165	6.00E-57	Medicago truncatula	clone mth2-32m22, WORKING DRAFT SEQUENC
89	G482	BQ104671	2.00E-55	Rosa hybrid cultivar	fc0546.e Rose Petals (Fragrant Cloud)
89	G482	BI469382	4.00E-55	Glycine max	sai11b10.y1 Gm-c1053 Glycine max cDNA clone GEN
89	G482	AAAA01003638	1.00E-54	Oryza sativa (indica cultivar-group)	( ) scaffold003638
89	G482	AP005193	1.00E-54	Oryza sativa (japonica cultivar-group)	( ) chromosome 7 clo
89	G482	BU880488	1.00E-53	Populus balsamifera subsp. trichocarpa	UM49TG09 Populus flo
89	G482	BJ248969	2.00E-53	Triticum aestivum	BJ248969 Y. Ogihara unpublished cDNA libr
89	G482	AC120529	4.00E-53	Oryza sativa	chromosome 3 clone OSJNBa0039N21, *** SEQUENCI
89	G482	BU896236	7.00E-53	Populus tremula x Populus tremuloides	X037F04 Populus wood
89	G482	gi115840	1.40E-46	Zea mays	CCAAT-BINDING TRANSCRIPTION FACTOR SUBUNIT A (CB
89	G482	gi20160792	2.30E-41	Oryza sativa (japonica cultivar-group)	putative CAAT-box
93	G489	BH679015	1.00E-111	Brassica oleracea	BOHXO96TF BO_2_3_KB Brassica oleracea gen
93	G489	AC136503	1.00E-75	Medicago truncatula	clone mth2-15n1, WORKING DRAFT SEQUENCE
93	G489	BQ118033	8.00E-73	Solanum tuberosum	EST603609 mixed potato tissues Solanum tu
93	G489	BU873518	4.00E-68	Populus balsamifera subsp. trichocarpa	Q056D09 Populus flow
93	G489	BI934205	2.00E-67	Lycopersicon	EST554094 tomato flower,

				esculentum	anthesis L
93	G489	BQ797616	1.00E-66	Vitis vinifera	EST 6554 Ripening Grape berries Lambda Zap I
93	G489	BM064398	4.00E-63	Capsicum annuum	KS01066E11 KS01 Capsicum annuum cDNA, mRNA
93	G489	BU927107	4.00E-60	Glycine max	sas95f12.y1 Gm-c1036 Glycine max cDNA clone SOY
93	G489	BQ993879	6.00E-59	Lactuca sativa	QGF5L12.yg.ab1 QG_EFGHJ lettuce serriola Lac
93	G489	AP004113	1.00E-58	Oryza sativa	chromosome 2 clone OJ1116_A06, *** SEQUENCING
93	G489	gi5257260	6.20E-46	Oryza sativa	Similar to sequence of BAC F7G19 from Arabid
93	G489	gi20804442	6.60E-19	Oryza sativa (japonica cultivar-group)	hypothetical prote
93	G489	gi18481626	3.90E-09	Zea mays	repressor protein.
93	G489	gi1808688	0.051	Sporobolus stapfianus	hypothetical protein.
93	G489	gi8096192	0.21	Lilium longiflorum	gH2A.1.
93	G489	gi2130105	0.25	Triticum aestivum	histone H2A.4 - wheat.
93	G489	gi297871	0.27	Picea abies	histone H2A.
93	G489	gi297887	0.31	Daucus carota	glycine rich protein.
93	G489	gi15214035	0.75	Cicer arietinum	HISTONE H2A.
93	G489	gi2317760	0.75	Pinus taeda	H2A homolog.
127	G634	OSGT2	2.00E-47	Oryza sativa	O.sativa gt-2 gene.
127	G634	BU049946	1.00E-46	Zea mays	1111017E09.y1 1111 - Unigene III from Maize Genome
127	G634	AF372499	6.00E-38	Glycine max	GT-2 factor mRNA, partial cds.
127	G634	AB052729	4.00E-37	Pisum sativum	mRNA for DNA-binding protein DF1, complete cd
127	G634	BU889446	4.00E-36	Populus tremula	P021A05 Populus petioles cDNA library Popul
127	G634	BH436958	2.00E-35	Brassica oleracea	BOHBE67TF BOHB Brassica oleracea genomic
127	G634	AI777252	3.00E-35	Lycopersicon esculentum	EST258217 tomato resistant, Cornell
127	G634	AW686754	1.00E-33	Medicago truncatula	NF042C08NR1F1000 Nodulated root Medicag
127	G634	AV410715	4.00E-33	Lotus japonicus	AV410715 Lotus japonicus young plants (two-
127	G634	AI730933	8.00E-30	Gossypium hirsutum	BNLGHi8208 Six-day Cotton fiber Gossypiu
127	G634	gi13786451	3.20E-78	Oryza sativa	putative transcription factor.
127	G634	gi13646986	3.50E-66	Pisum sativum	DNA-binding protein DF1.
127	G634	gi18182311	2.70E-38	Glycine max	GT-2 factor.
127	G634	gi20161567	8.90E-11	Oryza sativa (japonica cultivar-	hypothetical prote

				group)	
127	G634	gi170271	4.70E-08	Nicotiana tabacum	DNA-binding protein.
127	G634	gi18349	0.0027	Daucus carota	glycine rich protein (AA 1 - 96).
127	G634	gi21388658	0.027	Physcomitrella patens	glycine-rich RNA binding protein.
127	G634	gi21322752	0.052	Triticum aestivum	cold shock protein-1.
127	G634	gi3126963	0.057	Elaeagnus umbellata	acidic chitinase.
127	G634	gi1166450	0.087	Lycopersicon esculentum	Tfm5.
147	G682	BU831849	8.00E-25	Populus tremula x Populus tremuloides	T026E01 Populus apica
147	G682	BU872107	8.00E-25	Populus balsamifera subsp. trichocarpa	Q039C07 Populus flow
147	G682	BM437313	1.00E-20	Vitis vinifera	VVA017F06_54121 An expressed sequence tag da
147	G682	BI699876	4.00E-19	Glycine max	sag49b09.y1 Gm-c1081 Glycine max cDNA clone GEN
147	G682	BH961028	1.00E-16	Brassica oleracea	odj30d06.g1 B.oleracea002 Brassica olerac
147	G682	AL750151	2.00E-14	Pinus pinaster	AL750151 AS Pinus pinaster cDNA clone AS06C1
147	G682	BJ476463	1.00E-13	Hordeum vulgare subsp. vulgare	BJ476463 K. Sato unpublished
147	G682	AJ485557	1.00E-13	Hordeum vulgare	AJ485557 S00011 Hordeum vulgare cDNA clone
147	G682	CA762299	2.00E-13	Oryza sativa (indica cultivar-group)	BR060003B10F03.ab1 IRR
147	G682	CA736777	2.00E-12	Triticum aestivum	wpi1s.pk008.n12 wpi1s Triticum aestivum c
147	G682	gi23476287	8.30E-12	Gossypium hirsutum	myb-like transcription factor 2.
147	G682	gi23476291	8.30E-12	Gossypium raimondii	myb-like transcription factor 2.
147	G682	gi23476293	8.30E-12	Gossypium herbaceum	myb-like transcription factor 2.
147	G682	gi23476295	8.30E-12	Gossypioides kirkii	myb-like transcription factor 2.
147	G682	gi15042120	2.20E-11	Zea luxurians	CI protein.
147	G682	gi19548449	2.20E-11	Zea mays	P-type R2R3 Myb protein.
147	G682	gi9954118	2.80E-11	Solanum tuberosum	tuber-specific and sucrose-responsive e
147	G682	gi15042108	4.60E-11	Zea mays subsp. parviglumis	CI protein.
147	G682	gi15082210	1.50E-10	Fragaria x ananassa	transcription factor MYB1.
147	G682	gi22266669	1.50E-10	Vitis labrusca x Vitis vinifera	myb-related transcription
167	G864	BH472654	1.00E-105	Brassica oleracea	BOHPF07TF BOHP Brassica oleracea genomic
167	G864	AP004902	2.00E-44	Lotus japonicus	genomic DNA, chromosome 2, clone:LjT04G24,
167	G864	BM886518	5.00E-40	Glycine max	sam17f08.y1 Gm-c1068

					Glycine max cDNA clone SOY
167	G864	AW685524	5.00E-39	Medicago truncatula	NF031C12NR1F1000 Nodulated root Medicag
167	G864	AP001800	6.00E-36	Oryza sativa	genomic DNA, chromosome 1, PAC clone:P0443E05.
167	G864	LEU89257	6.00E-32	Lycopersicon esculentum	DNA-binding protein Pti6 mRNA, comp
167	G864	AAAA01000263	7.00E-31	Oryza sativa (indica cultivar-group)	( ) scaffold000263
167	G864	BQ873772	8.00E-30	Lactuca sativa	QGI2I03.yg.ab1 QG_ABCDI lettuce salinas Lact
167	G864	AF058827	7.00E-29	Nicotiana tabacum	TSI1 (Tsi1) mRNA, complete cds.
167	G864	BZ419846	3.00E-25	Zea mays	if61a07.b1 WGS-ZmaysF (DH5a methyl filtered) Zea m
167	G864	gi8096469	1.60E-38	Oryza sativa	Similar to *Arabidopsis thaliana* chromosome 4
167	G864	gi2213785	1.00E-34	Lycopersicon esculentum	Pti6.
167	G864	gi23617235	3.70E-25	Oryza sativa (japonica cultivar-group)	contains ESTs AU16
167	G864	gi3065895	7.60E-25	Nicotiana tabacum	TSI1.
167	G864	gi3264767	1.90E-21	Prunus armeniaca	AP2 domain containing protein.
167	G864	gi8571476	4.30E-21	Atriplex hortensis	apetala2 domain-containing protein.
167	G864	gi17385636	2.80E-20	Matricaria chamomilla	ethylene-responsive element binding
167	G864	gi8809571	4.50E-20	Nicotiana sylvestris	ethylene-responsive element binding
167	G864	gi7528276	5.70E-20	Mesembryanthemum crystallinum	AP2-related transcription f
167	G864	gi21908036	9.30E-20	Zea mays	DRE binding factor 1.
169	G867	BQ971511	2.00E-94	Helianthus annuus	QHB7E05.yg.ab1 QH_ABCDI sunflower RHA801
169	G867	AP003450	6.00E-85	Oryza sativa	chromosome 1 clone P0034C09, *** SEQUENCING IN
169	G867	AC135925	1.00E-80	Oryza sativa (japonica cultivar-group)	( ) chromosome 5 clo
169	G867	AAAA01000997	1.00E-79	Oryza sativa (indica cultivar-group)	( ) scaffold000997
169	G867	BQ405698	2.00E-77	Gossypium arboreum	GA__Ed0085H02f Gossypium arboreum 7-10 d
169	G867	BZ015521	4.00E-69	Brassica oleracea	oeg86a05.g1 B.oleracea002 Brassica olerac
169	G867	BF520598	2.00E-66	Medicago truncatula	EST458071 DSIL Medicago truncatula cDNA

169	G867	BU994579	4.00E-64	Hordeum vulgare subsp. vulgare	HM07I08r HM Hordeum vulgare
169	G867	BF424857	2.00E-62	Glycine max	su59h03.y1 Gm-c1069 Glycine max cDNA clone GENO
169	G867	BU871082	1.00E-61	Populus balsamifera subsp. trichocarpa	Q026F06 Populus flow
169	G867	gi18565433	2.40E-85	Oryza sativa (japonica cultivar-group)	DNA-binding protei
169	G867	gi12328560	2.90E-73	Oryza sativa	putative DNA binding protein RAV2.
169	G867	gi10798644	7.30E-13	Nicotiana tabacum	AP2 domain-containing transcription fac
169	G867	gi18266198	2.50E-10	Narcissus pseudonarcissus	AP-2 domain containing protein.
169	G867	gi20340233	2.50E-10	Thellungiella halophila	ethylene responsive element bindi
169	G867	gi22074046	1.50E-09	Lycopersicon esculentum	transcription factor JERF1.
169	G867	gi3264767	6.90E-09	Prunus armeniaca	AP2 domain containing protein.
169	G867	gi18496063	7.10E-09	Fagus sylvatica	ethylene responsive element binding prote
169	G867	gi13173164	8.30E-09	Pisum sativum	APETAL2-like protein.
169	G867	gi1730475	8.70E-09	Hordeum vulgare	viviparous-1.
185	G912	BH498662	2.00E-93	Brassica oleracea	BOGTO66TR BOGT Brassica oleracea genomic
185	G912	AF084185	2.00E-75	Brassica napus	dehydration responsive element binding prote
185	G912	AF211531	1.00E-59	Nicotiana tabacum	Avr9/Cf-9 rapidly elicited protein 111B
185	G912	AY034473	1.00E-55	Lycopersicon esculentum	putative transcriptional activator
185	G912	BG321601	4.00E-53	Descurainia sophia	Ds01_01h03_R Ds01_AAFC_ECORC_cold_ stress
185	G912	AB080965	9.00E-53	Prunus avium	DREB1-like gene for dehydratiion responsive el
185	G912	BG590659	4.00E-51	Solanum tuberosum	EST498501 P. infestans-challenged leaf So
185	G912	BG644969	1.00E-50	Medicago truncatula	EST506588 KV3 Medicago truncatula cDNA
185	G912	BU016783	2.00E-49	Helianthus annuus	QHE14A02.yg.ab1 QH_EFGHJ sunflower RHA280
185	G912	BU871514	1.00E-47	Populus balsamifera subsp. trichocarpa	Q031D09 Populus flow
185	G912	gi5616086	5.90E-73	Brassica napus	dehydration responsive element binding pro
185	G912	gi12003384	5.20E-58	Nicotiana tabacum	Avr9/Cf-9 rapidly elicited protein 111B
185	G912	gi23495458	3.90E-53	Prunus avium	dehydratiion responsive

					element binding prot
185	G912	gi18535580	2.00E-49	Lycopersicon esculentum	putative transcriptional activato
185	G912	gi19071243	1.30E-45	Hordeum vulgare	CRT/DRE binding factor 1.
185	G912	gi24474328	8.20E-44	Oryza sativa (japonica cultivar-group)	apetala2 domain-co
185	G912	gi6983877	9.00E-38	Oryza sativa	Similar to mRNA for DREB1A (AB007787).
185	G912	gi17148651	3.90E-35	Secale cereale	CBF-like protein.
185	G912	gi20152903	1.40E-32	Hordeum vulgare subsp. vulgare	CRT/DRE binding factor 2.
185	G912	gi17226801	2.10E-31	Triticum aestivum	putative CRT/DRE-binding factor.
187	G913	AI352878	4.00E-87	Brassica napus	MB72-11D PZ204.BNlib Brassica napus cDNA clo
187	G913	BH536782	1.00E-59	Brassica oleracea	BOGCX29TR BOGC Brassica oleracea genomic
187	G913	AW033835	2.00E-46	Lycopersicon esculentum	EST277406 tomato callus, TAMU Lycop
187	G913	BQ411166	1.00E-43	Gossypium arboreum	GA_Ed0037B05f Gossypium arboreum 7-10 d
187	G913	BQ165313	5.00E-43	Medicago truncatula	EST611182 KVKC Medicago truncatula cDNA
187	G913	AP006060	5.00E-43	Oryza sativa (japonica cultivar-group)	( ) chromosome 2 clo
187	G913	AAAA01000810	2.00E-42	Oryza sativa (indica cultivar-group)	( ) scaffold000810
187	G913	OSJN00128	2.00E-38	Oryza sativa	chromosome 4 clone OSJNBA0088I22, *** SEQUENC
187	G913	BQ976989	3.00E-31	Helianthus annuus	QHI23I22.yg.ab1 QH_ABCDI sunflower RHA801
187	G913	BQ592028	6.00E-30	Beta vulgaris	E012695-024-021-K17-SP6 MPIZ-ADIS-024-develop
187	G913	gi14140155	1.60E-32	Oryza sativa	putative AP2 domain transcription factor.
187	G913	gi12003382	1.40E-30	Nicotiana tabacum	Avr9/Cf-9 rapidly elicited protein 111A
187	G913	gi20303570	1.40E-30	Oryza sativa (japonica cultivar-group)	putative transcrip
187	G913	gi18535580	3.80E-30	Lycopersicon esculentum	putative transcriptional activato
187	G913	gi23495460	4.40E-29	Prunus avium	dehydration responsive element binding prote
187	G913	gi5616086	6.50E-28	Brassica napus	dehydration responsive element binding pro
187	G913	gi21908034	1.40E-25	Zea mays	DRE binding factor 2.
187	G913	gi19071243	1.20E-21	Hordeum vulgare	CRT/DRE binding factor 1.
187	G913	gi17148649	2.30E-17	Secale cereale	CBF-like protein.

187	G913	gi8571476	2.30E-17	Atriplex hortensis	apetala2 domain-containing protein.
189	G922	AP004485	1.0e-999	Lotus japonicus	genomic DNA, chromosome 2, clone:LjT08D14,
189	G922	AP003259	1.00E-130	Oryza sativa	chromosome 1 clone P0466H10, *** SEQUENCING IN
189	G922	AAAA01000374	1.00E-130	Oryza sativa (indica cultivar-group)	( ) scaffold000374
189	G922	BH493536	1.00E-121	Brassica oleracea	BOGXB10TR BOGX Brassica oleracea genomic
189	G922	CNS08CCP	1.00E-92	Oryza sativa (japonica cultivar-group)	( ) chromosome 12 cl
189	G922	BG643567	6.00E-82	Lycopersicon esculentum	EST511761 tomato shoot/meristem Lyc
189	G922	BQ124898	2.00E-81	Medicago truncatula	EST610474 GLSD Medicago truncatula cDNA
189	G922	BU764181	2.00E-71	Glycine max	sas53f07.y1 Gm-c1023 Glycine max cDNA clone SOY
189	G922	BG595716	3.00E-62	Solanum tuberosum	EST494394 cSTS Solanum tuberosum cDNA clo
189	G922	AF378125	6.00E-55	Vitis vinifera	GAI-like protein 1 (GAI1) gene, complete cds
189	G922	gi22830925	6.30E-127	Oryza sativa (japonica cultivar-group)	putative gibberell
189	G922	gi13365610	3.00E-57	Pisum sativum	SCARECROW.
189	G922	gi13170126	5.20E-55	Brassica napus	unnamed protein product.
189	G922	gi10178637	6.30E-51	Zea mays	SCARECROW.
189	G922	gi13937306	2.30E-50	Oryza sativa	gibberellin-insensitive protein OsGAI.
189	G922	gi18254373	9.20E-50	Hordeum vulgare	nuclear transcription factor SLN1.
189	G922	gi5640157	2.60E-49	Triticum aestivum	gibberellin response modulator.
189	G922	gi20257451	3.10E-49	Calycadenia multiglandulosa	GIA/RGA-like gibberellin resp
189	G922	gi13620224	1.30E-46	Lycopersicon esculentum	lateral suppressor.
189	G922	gi13620166	2.20E-41	Capsella rubella	hypothetical protein.
191	G926	BU573158	1.00E-56	Prunus dulcis	PA__Ea0003A12f Almond developing seed Prunus
191	G926	BI310587	2.00E-55	Medicago truncatula	EST5312337 GESD Medicago truncatula cDN
191	G926	BQ624240	1.00E-47	Citrus sinensis	USDA-FP_01331 Ridge pineapple sweet orange
191	G926	BH443554	3.00E-44	Brassica oleracea	BOHGN12TR BOHG Brassica oleracea genomic
191	G926	BNU33884	2.00E-39	Brassica napus	clone bncbf-b1 CCAAT-binding factor B subuni
191	G926	BF113081	8.00E-38	Lycopersicon	EST440591 tomato breaker

				esculentum	fruit Lyco
191	G926	BG886494	2.00E-36	Solanum tuberosum	EST512345 cSTD Solanum tuberosum cDNA clo
191	G926	AW472517	3.00E-36	Glycine max	si26c12.y1 Gm-r1030 Glycine max cDNA clone GENO
191	G926	BQ407583	6.00E-36	Gossypium arboreum	GA_Ed0108F07f Gossypium arboreum 7-10 d
191	G926	BG343051	7.00E-34	Hordeum vulgare	HVSMEg0001N16f Hordeum vulgare pre-anthesis
191	G926	gi1173616	9.70E-41	Brassica napus	CCAAT-binding factor B subunit homolog.
191	G926	gi2826786	1.10E-27	Oryza sativa	RAPB protein.
191	G926	gi7141243	5.80E-27	Vitis riparia	transcription factor.
191	G926	gi4731314	4.00E-19	Nicotiana tabacum	CCAAT-binding transcription factor subu
191	G926	gi2104675	0.0061	Vicia faba	transcription factor.
191	G926	gi21667471	0.64	Hordeum vulgare	CONSTANS-like protein.
191	G926	gi13775107	0.67	Phaseolus vulgaris	bZIP transcription factor 2.
191	G926	gi1096930	0.69	Solanum tuberosum	H ATPase inhibitor.
191	G926	gi24413952	0.72	Oryza sativa (japonica cultivar-group)	putative iron supe
191	G926	gi1839593	0.78	Zea mays	heat shock protein 70 homolog {clone CHEM 3} [Ze
199	G975	BH477624	1.00E-69	Brassica oleracea	BOGNB10TF BOGN Brassica oleracea genomic
199	G975	CA486875	3.00E-64	Triticum aestivum	WHE4337_A02_A03ZS Wheat meiotic anther cD
199	G975	BI978981	2.00E-60	Rosa chinensis	zD09 Old Blush petal SMART library Rosa chin
199	G975	AP004869	9.00E-60	Oryza sativa (japonica cultivar-group)	( ) chromosome 2 clo
199	G975	BU978490	1.00E-58	Hordeum vulgare subsp. vulgare	HA13G05r HA Hordeum vulgare
199	G975	BG642554	8.00E-57	Lycopersicon esculentum	EST356031 tomato flower buds, anthe
199	G975	BI958226	2.00E-54	Hordeum vulgare	HVSMEn0013P17f Hordeum vulgare rachis EST 1
199	G975	BQ104740	1.00E-52	Rosa hybrid cultivar	fc0212.e Rose Petals (Fragrant Cloud)
199	G975	AW705973	3.00E-51	Glycine max	sk64c02.y1 Gm-c1016 Glycine max cDNA clone GENO
199	G975	AP003615	1.00E-47	Oryza sativa	chromosome 6 clone P0486H12, *** SEQUENCING IN
199	G975	gi18650662	1.80E-25	Lycopersicon esculentum	ethylene response factor 1.

199	G975	gi131754	2.10E-22	Lupinus polyphyllus	PPLZ02 PROTEIN.
199	G975	gi3065895	9.20E-20	Nicotiana tabacum	TSI1.
199	G975	gi8571476	9.30E-20	Atriplex hortensis	apetala2 domain-containing protein.
199	G975	gi19920190	1.90E-19	Oryza sativa (japonica cultivar-group)	Putative AP2 domai
199	G975	gi21908036	8.40E-19	Zea mays	DRE binding factor 1.
199	G975	gi4099914	1.10E-18	Stylosanthes hamata	ethylene-responsive element binding p
199	G975	gi10567106	1.60E-18	Oryza sativa	osERF3.
199	G975	gi8809573	9.60E-18	Nicotiana sylvestris	ethylene-responsive element binding
199	G975	gi7528276	1.20E-17	Mesembryanthemum crystallinum	AP2-related transcription f
221	G1069	BZ025139	1.00E-111	Brassica oleracea	oeh63d12.g1 B.oleracea002 Brassica olerac
221	G1069	AP004971	1.00E-93	Lotus japonicus	genomic DNA, chromosome 5, clone:LjT45G21,
221	G1069	AP004020	2.00E-79	Oryza sativa	chromosome 2 clone OJ1119_A01, *** SEQUENCING
221	G1069	AAAA01017331	2.00E-70	Oryza sativa (indica cultivar-group)	( ) scaffold017331
221	G1069	BQ165495	2.00E-62	Medicago truncatula	EST611364 KVKC Medicago truncatula cDNA
221	G1069	AC135209	2.00E-61	Oryza sativa (japonica cultivar-group)	( ) chromosome 3 clo
221	G1069	AW621455	4.00E-59	Lycopersicon esculentum	EST312253 tomato root during/after
221	G1069	BM110212	4.00E-58	Solanum tuberosum	EST557748 potato roots Solanum tuberosum
221	G1069	BQ785950	7.00E-58	Glycine max	saq61f09.y1 Gm-c1076 Glycine max cDNA clone SOY
221	G1069	BQ863249	1.00E-57	Lactuca sativa	QGC23G02.yg.ab1 QG_ABCDI lettuce salinas Lac
221	G1069	gi24059979	2.10E-38	Oryza sativa (japonica cultivar-group)	similar to DNA-bin
221	G1069	gi15528814	4.50E-36	Oryza sativa	hypothetical protein~similar to *Arabidopsis*
221	G1069	gi4165183	7.60E-25	Antirrhinum majus	SAP1 protein.
221	G1069	gi2213534	1.20E-19	Pisum sativum	DNA-binding PD1-like protein.
221	G1069	gi2459999	1	Chlamydomonas reinhardtii	tubulin Uni3.
221	G1069	gi100872	1	Zea mays	MFS18 protein - maize.
221	G1069	gi1362165	1	Hordeum vulgare	hypothetical protein 2 (clone ES1A) - bar
223	G1073	AAAA01000486	4.00E-74	Oryza sativa (indica	( ) scaffold000486

				cultivar-group)	
223	G1073	AP004165	4.00E-74	Oryza sativa	chromosome 2 clone OJ1479_B12, *** SEQUENCING
223	G1073	AP005477	2.00E-67	Oryza sativa (japonica cultivar-group)	( ) chromosome 6 clo
223	G1073	BZ412041	3.00E-65	Zea mays	OGACG56TC ZM_0.7_1.5_KB Zea mays genomic clone ZMM
223	G1073	AJ502190	3.00E-64	Medicago truncatula	AJ502190 MTAMP Medicago truncatula cDNA
223	G1073	BQ865858	4.00E-63	Lactuca sativa	QGC6B08.yg.ab1 QG_ABCDI lettuce salinas Lact
223	G1073	BH975957	5.00E-63	Brassica oleracea	odh67e11.g1 B.oleracea002 Brassica olerac
223	G1073	BG134451	8.00E-62	Lycopersicon esculentum	EST467343 tomato crown gall Lycoper
223	G1073	AP004971	3.00E-60	Lotus japonicus	genomic DNA, chromosome 5, clone:LjT45G21,
223	G1073	BM110212	7.00E-58	Solanum tuberosum	EST557748 potato roots Solanum tuberosum
223	G1073	gi15528814	5.50E-38	Oryza sativa	hypothetical protein~similar to *Arabidopsis*
223	G1073	gi24059979	1.30E-29	Oryza sativa (japonica cultivar-group)	similar to DNA-bin
223	G1073	gi2213536	1.20E-21	Pisum sativum	DNA-binding protein PD1.
223	G1073	gi4165183	5.70E-20	Antirrhinum majus	SAP1 protein.
223	G1073	gi1166450	0.00059	Lycopersicon esculentum	Tfm5.
223	G1073	gi11545668	0.0051	Chlamydomonas reinhardtii	CIA5.
223	G1073	gi4755087	0.0054	Zea mays	aluminum-induced protein; Al-induced protein.
223	G1073	gi395147	0.0068	Nicotiana tabacum	glycine-rich protein.
223	G1073	gi21068672	0.017	Cicer arietinum	putative glicine-rich protein.
223	G1073	gi1346181	0.017	Sinapis alba	GLYCINE-RICH RNA-BINDING PROTEIN GRP2A.
225	G1075	BH596283	1.00E-108	Brassica oleracea	BOGBL42TR BOGB Brassica oleracea genomic
225	G1075	BQ165495	5.00E-88	Medicago truncatula	EST611364 KVKC Medicago truncatula cDNA
225	G1075	AAAA01003389	3.00E-84	Oryza sativa (indica cultivar-group)	( ) scaffold003389
225	G1075	OSJN00182	3.00E-84	Oryza sativa	chromosome 4 clone OSJNBa0086O06, *** SEQUENC
225	G1075	BZ412041	1.00E-76	Zea mays	OGACG56TC ZM_0.7_1.5_KB Zea mays genomic clone ZMM

225	G1075	AP005653	1.00E-68	Oryza sativa (japonica cultivar-group)	( ) chromosome 2 clo
225	G1075	BQ863249	3.00E-65	Lactuca sativa	QGC23G02.yg.ab1 QG_ABCD1 lettuce salinas Lac
225	G1075	BM110212	2.00E-63	Solanum tuberosum	EST557748 potato roots Solanum tuberosum
225	G1075	BQ838600	8.00E-63	Triticum aestivum	WHE2912_D12_H24ZS Wheat aluminum-stressed
225	G1075	AP004971	4.00E-62	Lotus japonicus	genomic DNA, chromosome 5, clone:LjT45G21,
225	G1075	gi15528814	3.80E-39	Oryza sativa	hypothetical protein~similar to *Arabidopsis*
225	G1075	gi24059979	6.60E-35	Oryza sativa (japonica cultivar-group)	similar to DNA-bin
225	G1075	gi4165183	7.30E-20	Antirrhinum majus	SAP1 protein.
225	G1075	gi2213534	2.50E-19	Pisum sativum	DNA-binding PD1-like protein.
225	G1075	gi3810890	3.70E-05	Cucumis sativus	glycine-rich protein-2.
225	G1075	gi7489009	0.0001	Lycopersicon esculentum	glycine-rich protein (clone w10-1
225	G1075	gi4115615	0.0018	Zea mays	root cap-specific glycine-rich protein.
225	G1075	gi1628463	0.004	Silene latifolia	Men-4.
225	G1075	gi395147	0.005	Nicotiana tabacum	glycine-rich protein.
225	G1075	gi121631	0.0056	Nicotiana sylvestris	GLYCINE-RICH CELL WALL STRUCTURAL PR
269	G1411	BZ017225	3.00E-51	Brassica oleracea	oei67e03.b1 B.oleracea002 Brassica olerac
269	G1411	BQ138607	8.00E-44	Medicago truncatula	NF005C01PH1F1004 Phoma-infected Medicag
269	G1411	BQ786702	5.00E-36	Glycine max	saq72b07.y1 Gm-c1076 Glycine max cDNA clone SOY
269	G1411	BM062508	7.00E-32	Capsicum annuum	KS01043F09 KS01 Capsicum annuum cDNA, mRNA
269	G1411	AAAA01000832	2.00E-30	Oryza sativa (indica cultivar-group)	( ) scaffold000832
269	G1411	OSJN00240	2.00E-30	Oryza sativa	genomic DNA, chromosome 4, BAC clone: OSJNBa0
269	G1411	BE419451	2.00E-29	Triticum aestivum	WWS012.C2R000101 ITEC WWS Wheat Scutellum
269	G1411	CA014817	6.00E-29	Hordeum vulgare subsp. vulgare	HT12H01r HT Hordeum vulgare
269	G1411	BE642320	1.00E-28	Ceratopteris richardii	Cri2_5_L17_SP6 Ceratopteris Spore Li
269	G1411	BE494041	2.00E-27	Secale cereale	WHE1277_B09_D17ZS Secale cereale anther cDNA
269	G1411	gi20160854	1.40E-29	Oryza sativa (japonica cultivar-	hypothetical prote

				group)	
269	G1411	gi14140141	1.50E-24	Oryza sativa	putative AP2-related transcription factor.
269	G1411	gi3342211	1.40E-23	Lycopersicon esculentum	Pti4.
269	G1411	gi10798644	2.30E-23	Nicotiana tabacum	AP2 domain-containing transcription fac
269	G1411	gi8809571	2.30E-23	Nicotiana sylvestris	ethylene-responsive element binding
269	G1411	gi24817250	3.00E-23	Cicer arietinum	transcription factor EREBP-like protein.
269	G1411	gi3264767	3.00E-23	Prunus armeniaca	AP2 domain containing protein.
269	G1411	gi1688233	3.80E-23	Solanum tuberosum	DNA binding protein homolog.
269	G1411	gi7528276	3.80E-23	Mesembryanthemum crystallinum	AP2-related transcription f
269	G1411	gi21304712	6.20E-23	Glycine max	ethylene-responsive element binding protein 1
277	G1451	AB071298	1.0e-999	Oryza sativa	OsARF8 mRNA for auxin response factor 8, parti
277	G1451	AY105215	1.00E-157	Zea mays	PCO121637 mRNA sequence.
277	G1451	AW690130	1.00E-109	Medicago truncatula	NF028B12ST1F1000 Developing stem Medica
277	G1451	BQ862285	1.00E-108	Lactuca sativa	QGC20K23.yg.ab1 QG_ABCDI lettuce salinas Lac
277	G1451	BG597435	1.00E-107	Solanum tuberosum	EST496113 cSTS Solanum tuberosum cDNA clo
277	G1451	BJ303602	1.00E-104	Triticum aestivum	BJ303602 Y. Ogihara unpublished cDNA libr
277	G1451	OSA306306	1.00E-103	Oryza sativa (japonica cultivar-group)	Oryza sativa subsp.
277	G1451	BQ595269	1.00E-89	Beta vulgaris	E012710-024-023-D13-SP6 MPIZ-ADIS-024-develop
277	G1451	CA801218	1.00E-86	Glycine max	sau02f06.y2 Gm-c1062 Glycine max cDNA clone SOY
277	G1451	BG159611	8.00E-79	Sorghum bicolor	OV2_6_G07.b1_A002 Ovary 2 (OV2) Sorghum bic
277	G1451	gi19352049	3.70E-247	Oryza sativa	auxin response factor 8.
277	G1451	gi20805236	3.10E-126	Oryza sativa (japonica cultivar-group)	auxin response fac
277	G1451	gi24785191	4.10E-55	Nicotiana tabacum	hypothetical protein.
277	G1451	gi23343944	2.40E-28	Mirabilis jalapa	auxin-responsive factor protein.
277	G1451	gi20269053	7.00E-10	Populus tremula x Populus tremuloides	aux/IAA protein.
277	G1451	gi287566	3.10E-06	Vigna radiata	ORF.
277	G1451	gi114733	1.10E-05	Glycine max	AUXIN-INDUCED

					PROTEIN AUX22.
277	G1451	gi871511	2.40E-05	Pisum sativum	auxin-induced protein.
277	G1451	gi18697008	0.00027	Zea mays	unnamed protein product.
277	G1451	gi17976835	0.00068	Pinus pinaster	putative auxin induced transcription facto
303	G1543	AF145727	4.00E-51	Oryza sativa	homeodomain leucine zipper protein (hox3) mRNA
303	G1543	CA030381	6.00E-41	Hordeum vulgare subsp. vulgare	HX06O07r HX Hordeum vulgare
303	G1543	BQ741095	6.00E-39	Glycine max	saq14c10.y1 Gm-c1045 Glycine max cDNA clone SOY
303	G1543	AT002118	1.00E-38	Brassica rapa subsp. pekinensis	AT002118 Flower bud cDNA Br
303	G1543	BQ857226	2.00E-37	Lactuca sativa	QGB6P03.yg.ab1 QG_ABCDI lettuce salinas Lact
303	G1543	AB028075	4.00E-37	Physcomitrella patens	mRNA for homeobox protein PpHB4, comp
303	G1543	PBPHZ4GEN	4.00E-37	Pimpinella brachycarpa	P.brachycarpa mRNA for homeobox-leu
303	G1543	LEHDZIPP	5.00E-37	Lycopersicon esculentum	L.esculentum mRNA for HD-ZIP protei
303	G1543	AF443619	1.00E-36	Craterostigma plantagineum	homeodomain leucine zipper prote
303	G1543	AJ498394	2.00E-36	Medicago truncatula	AJ498394 MTPOSE Medicago truncatula cDN
303	G1543	gi5006851	8.30E-51	Oryza sativa	homeodomain leucine zipper protein.
303	G1543	gi20161555	1.70E-50	Oryza sativa (japonica cultivar-group)	putative homeodoma
303	G1543	gi18034437	1.60E-38	Craterostigma plantagineum	homeodomain leucine zipper pro
303	G1543	gi1149535	4.30E-38	Pimpinella brachycarpa	homeobox-leucine zipper protein.
303	G1543	gi992598	1.20E-37	Lycopersicon esculentum	HD-ZIP protein.
303	G1543	gi7415620	1.50E-37	Physcomitrella patens	homeobox protein PpHB4.
303	G1543	gi1234900	3.10E-37	Glycine max	homeobox-leucine zipper protein.
303	G1543	gi3868847	1.90E-35	Ceratopteris richardii	CRHB10.
303	G1543	gi8919876	1.90E-35	Capsella rubella	hypothetical protein.
303	G1543	gi1032372	3.20E-35	Helianthus annuus	homeodomain protein.
331	G1792	AI776626	5.00E-35	Lycopersicon esculentum	EST257726 tomato resistant, Cornell
331	G1792	BQ045702	1.00E-32	Solanum tuberosum	EST594820 P. infestans-challenged potato
331	G1792	BM178875	7.00E-32	Glycine max	saj60f01.y1 Gm-c1072 Glycine max cDNA clone SOY
331	G1792	BF649790	1.00E-31	Medicago truncatula	NF084C07EC1F1052 Elicited cell culture

331	G1792	BZ020356	1.00E-30	Brassica oleracea	oeg04a10.g1 B.oleracea002 Brassica olerac
331	G1792	BZ337899	3.00E-30	Sorghum bicolor	ia91f11.b1 WGS-SbicolorF (JM107 adapted met
331	G1792	AC025907	3.00E-30	Oryza sativa	chromosome 10 clone nbxb0094K20, *** SEQUENCIN
331	G1792	AAAA01002491	3.00E-30	Oryza sativa (indica cultivar-group)	( ) scaffold002491
331	G1792	BZ359367	8.00E-30	Zea mays	id72f11.b1 WGS-ZmaysF (JM107 adapted methyl filter
331	G1792	AC137635	2.00E-27	Oryza sativa (japonica cultivar-group)	Genomic sequence for
331	G1792	gi23452024	4.00E-26	Lycopersicon esculentum	transcription factor TSRF1.
331	G1792	gi1732406	2.10E-25	Nicotiana tabacum	S25-XP1 DNA binding protein.
331	G1792	gi12597874	3.70E-25	Oryza sativa	putative ethylene-responsive element binding
331	G1792	gi7528276	7.60E-25	Mesembryanthemum crystallinum	AP2-related transcription f
331	G1792	gi24060081	1.30E-23	Oryza sativa (japonica cultivar-group)	putative ethylene
331	G1792	gi8980313	1.80E-23	Catharanthus roseus	AP2-domain DNA-binding protein.
331	G1792	gi8809571	1.80E-23	Nicotiana sylvestris	ethylene-responsive element binding
331	G1792	gi17385636	1.20E-21	Matricaria chamomilla	ethylene-responsive element binding
331	G1792	gi21304712	3.10E-21	Glycine max	ethylene-responsive element binding protein 1
331	G1792	gi8571476	1.10E-20	Atriplex hortensis	apetala2 domain-containing protein.
341	G1820	AW776719	1.00E-43	Medicago truncatula	EST335784 DSIL Medicago truncatula cDNA
341	G1820	BM065544	3.00E-40	Capsicum annuum	KS07004F12 KS07 Capsicum annuum cDNA, mRNA
341	G1820	BG591677	4.00E-40	Solanum tuberosum	EST499519 P. infestans-challenged leaf So
341	G1820	BI701620	1.00E-38	Glycine max	sai18a04.y1 Gm-c1053 Glycine max cDNA clone GEN
341	G1820	BQ411597	3.00E-37	Gossypium arboreum	GA__Ed0041B06f Gossypium arboreum 7-10 d
341	G1820	BE208917	6.00E-37	Citrus x paradisi	GF-FV-P3F5 Marsh grapefruit young flavedo
341	G1820	BH725354	1.00E-36	Brassica oleracea	BOHVO37TF BO_2_3_KB Brassica oleracea gen
341	G1820	AW093662	9.00E-36	Lycopersicon esculentum	EST286842 tomato mixed elicitor, BT

341	G1820	BU819346	4.00E-35	Populus tremula	UA42BPF01 Populus tremula cambium cDNA libr
341	G1820	AAAA01002977	3.00E-34	Oryza sativa (indica cultivar-group)	( ) scaffold002977
341	G1820	gi5257260	1.40E-34	Oryza sativa	Similar to sequence of BAC F7G19 from Arabid
341	G1820	gi20804442	1.70E-15	Oryza sativa (japonica cultivar-group)	hypothetical prote
341	G1820	gi18481626	6.30E-08	Zea mays	repressor protein.
341	G1820	gi297871	0.39	Picea abies	histone H2A.
341	G1820	gi297887	0.41	Daucus carota	glycine rich protein.
341	G1820	gi2130105	0.54	Triticum aestivum	histone H2A.4 - wheat.
341	G1820	gi6782438	0.74	Nicotiana glauca	glycine-rich protein.
341	G1820	gi15214035	0.98	Cicer arietinum	HISTONE H2A.
341	G1820	gi2317760	0.98	Pinus taeda	H2A homolog.
341	G1820	gi1173628	0.99	Phalaenopsis sp. SM9108	glycine-rich protein.
343	G1836	BI701620	7.00E-35	Glycine max	sai18a04.y1 Gm-c1053 Glycine max cDNA clone GEN
343	G1836	AW776719	2.00E-33	Medicago truncatula	EST335784 DSIL Medicago truncatula cDNA
343	G1836	BQ411597	2.00E-33	Gossypium arboreum	GA__Ed0041B06f Gossypium arboreum 7-10 d
343	G1836	BM065544	2.00E-32	Capsicum annuum	KS07004F12 KS07 Capsicum annuum cDNA, mRNA
343	G1836	BG591677	3.00E-31	Solanum tuberosum	EST499519 P. infestans-challenged leaf So
343	G1836	BU819346	6.00E-31	Populus tremula	UA42BPF01 Populus tremula cambium cDNA libr
343	G1836	BH725354	4.00E-30	Brassica oleracea	BOHVO37TF BO_2_3_KB Brassica oleracea gen
343	G1836	BE208917	6.00E-30	Citrus x paradisi	GF-FV-P3F5 Marsh grapefruit young flavedo
343	G1836	AAAA01024926	5.00E-29	Oryza sativa (indica cultivar-group)	( ) scaffold024926
343	G1836	AW093662	9.00E-29	Lycopersicon esculentum	EST286842 tomato mixed elicitor, BT
343	G1836	gi5257260	2.10E-29	Oryza sativa	Similar to sequence of BAC F7G19 from Arabid
343	G1836	gi20804442	6.30E-16	Oryza sativa (japonica cultivar-group)	hypothetical prote
343	G1836	gi18481626	2.00E-06	Zea mays	repressor protein.
343	G1836	gi18539425	0.84	Pinus sylvestris	putative malate dehydrogenase.
343	G1836	gi122084	1	Hordeum vulgare	Histone H3.
343	G1836	gi225348	1	Hordeum vulgare subsp. vulgare	histone H3.
369	G1930	BU025988	5.00E-88	Helianthus annuus	QHG12J17.yg.ab1 QH_EFGHJ sunflower

					RHA280
369	G1930	AP003450	8.00E-80	Oryza sativa	chromosome 1 clone P0034C09, *** SEQUENCING IN
369	G1930	AC135925	7.00E-79	Oryza sativa (japonica cultivar-group)	( ) chromosome 5 clo
369	G1930	AAAA01000997	3.00E-78	Oryza sativa (indica cultivar-group)	( ) scaffold000997
369	G1930	BU994579	1.00E-65	Hordeum vulgare subsp. vulgare	HM07I08r HM Hordeum vulgare
369	G1930	BQ405698	1.00E-65	Gossypium arboreum	GA__Ed0085H02f Gossypium arboreum 7-10 d
369	G1930	BF520598	1.00E-64	Medicago truncatula	EST458071 DSIL Medicago truncatula cDNA
369	G1930	BZ015521	1.00E-64	Brassica oleracea	oeg86a05.g1 B.oleracea002 Brassica olerac
369	G1930	BF424857	2.00E-58	Glycine max	su59h03.y1 Gm-c1069 Glycine max cDNA clone GENO
369	G1930	BU870896	1.00E-56	Populus balsamifera subsp. trichocarpa	Q019F06 Populus flow
369	G1930	gi18565433	4.10E-74	Oryza sativa (japonica cultivar-group)	DNA-binding protei
369	G1930	gi12328560	1.80E-71	Oryza sativa	putative DNA binding protein RAV2.
369	G1930	gi10798644	1.40E-13	Nicotiana tabacum	AP2 domain-containing transcription fac
369	G1930	gi20340233	5.10E-11	Thellungiella halophila	ethylene responsive element bindi
369	G1930	gi4099921	1.30E-10	Stylosanthes hamata	EREBP-3 homolog.
369	G1930	gi18496063	1.60E-10	Fagus sylvatica	ethylene responsive element binding prote
369	G1930	gi22074046	2.10E-10	Lycopersicon esculentum	transcription factor JERF1.
369	G1930	gi3264767	2.30E-10	Prunus armeniaca	AP2 domain containing protein.
369	G1930	gi18266198	1.10E-09	Narcissus pseudonarcissus	AP-2 domain containing protein.
369	G1930	gi24940524	1.10E-09	Triticum aestivum	ethylene response element binding prote
407	G2133	BH420519	1.00E-53	Brassica oleracea	BOGUH88TF BOGU Brassica oleracea genomic
407	G2133	BG543936	6.00E-43	Brassica rapa subsp. pekinensis	E1686 Chinese cabbage etiol
407	G2133	AU292603	2.00E-28	Zinnia elegans	AU292603 zinnia cultured mesophyll cell equa
407	G2133	BE320193	6.00E-24	Medicago truncatula	NF024B04RT1F1029 Developing root Medica
407	G2133	AP003346	3.00E-22	Oryza sativa	chromosome 1 clone P0434C04, *** SEQUENCING IN

407	G2133	AAAA01000718	3.00E-22	Oryza sativa (indica cultivar-group)	( ) scaffold000718
407	G2133	AC124836	6.00E-22	Oryza sativa (japonica cultivar-group)	( ) chromosome 5 clo
407	G2133	BZ403609	2.00E-20	Zea mays	OGABN17TM ZM_0.7_1.5_KB Zea mays genomic clone ZMM
407	G2133	BM985484	6.00E-19	Thellungiella halophila	10_C12_T Ath Thellungiella halophil
407	G2133	BM403179	3.00E-17	Selaginella lepidophylla	SLA012F10_35741 An expressed seque
407	G2133	gi20161239	6.90E-24	Oryza sativa (japonica cultivar-group)	hypothetical prote
407	G2133	gi8571476	6.00E-17	Atriplex hortensis	apetala2 domain-containing protein.
407	G2133	gi14140155	7.80E-16	Oryza sativa	putative AP2 domain transcription factor.
407	G2133	gi5616086	7.00E-15	Brassica napus	dehydration responsive element binding pro
407	G2133	gi21908034	8.90E-15	Zea mays	DRE binding factor 2.
407	G2133	gi19071243	6.30E-14	Hordeum vulgare	CRT/DRE binding factor 1.
407	G2133	gi18535580	2.10E-13	Lycopersicon esculentum	putative transcriptional activato
407	G2133	gi1208496	3.30E-13	Nicotiana tabacum	EREBP-3.
407	G2133	gi8980313	4.40E-13	Catharanthus roseus	AP2-domain DNA-binding protein.
407	G2133	gi15488459	2.20E-12	Triticum aestivum	AP2-containing protein.
417	G2153	BH566718	1.00E-127	Brassica oleracea	BOHCV23TR BOHC Brassica oleracea genomic
417	G2153	AP004971	2.00E-90	Lotus japonicus	genomic DNA, chromosome 5, clone:LjT45G21,
417	G2153	AP004020	1.00E-79	Oryza sativa	chromosome 2 clone OJ1119_A01, *** SEQUENCING
417	G2153	AAAA01017331	2.00E-72	Oryza sativa (indica cultivar-group)	( ) scaffold017331
417	G2153	BQ165495	2.00E-67	Medicago truncatula	EST611364 KVKC Medicago truncatula cDNA
417	G2153	AP005653	1.00E-66	Oryza sativa (japonica cultivar-group)	( ) chromosome 2 clo
417	G2153	BQ785950	8.00E-64	Glycine max	saq61f09.y1 Gm-c1076 Glycine max cDNA clone SOY
417	G2153	BZ412041	3.00E-63	Zea mays	OGACG56TC ZM_0.7_1.5_KB Zea mays genomic clone ZMM
417	G2153	BM110212	3.00E-63	Solanum tuberosum	EST557748 potato roots Solanum tuberosum
417	G2153	BQ865858	7.00E-63	Lactuca sativa	QGC6B08.yg.ab1 QG_ABCDI lettuce salinas

					Lact
417	G2153	gi24059979	3.80E-39	Oryza sativa (japonica cultivar-group)	similar to DNA-bin
417	G2153	gi15528814	1.70E-36	Oryza sativa	hypothetical protein~similar to *Arabidopsis*
417	G2153	gi4165183	5.00E-21	Antirrhinum majus	SAP1 protein.
417	G2153	gi2213534	1.30E-19	Pisum sativum	DNA-binding PD1-like protein.
417	G2153	gi7439981	2.60E-08	Triticum aestivum	glycine-rich RNA-binding protein GRP1 -
417	G2153	gi21623	1.90E-06	Sorghum bicolor	glycine-rich RNA-binding protein.
417	G2153	gi11545668	3.50E-06	Chlamydomonas reinhardtii	CIA5.
417	G2153	gi21068672	6.60E-06	Cicer arietinum	putative glicine-rich protein.
417	G2153	gi7489714	6.60E-06	Zea mays	aluminum-induced protein al1 - maize.
417	G2153	gi395147	1.60E-05	Nicotiana tabacum	glycine-rich protein.
419	G2155	BG543096	2.00E-69	Brassica rapa subsp. pekinensis	E0571 Chinese cabbage etiol
419	G2155	BH480897	7.00E-66	Brassica oleracea	BOGRA01TF BOGR Brassica oleracea genomic
419	G2155	BG646893	2.00E-53	Medicago truncatula	EST508512 HOGA Medicago truncatula cDNA
419	G2155	BU023570	3.00E-44	Helianthus annuus	QHF11M19.yg.ab1 QH_EFGHJ sunflower RHA280
419	G2155	AP004020	2.00E-41	Oryza sativa	chromosome 2 clone OJ1119_A01, *** SEQUENCING
419	G2155	BI426899	4.00E-41	Glycine max	sag08g12.y1 Gm-c1080 Glycine max cDNA clone GEN
419	G2155	AAAA01000383	2.00E-40	Oryza sativa (indica cultivar-group)	( ) scaffold000383
419	G2155	AP004971	2.00E-40	Lotus japonicus	genomic DNA, chromosome 5, clone:LjT45G21,
419	G2155	AP005755	2.00E-40	Oryza sativa (japonica cultivar-group)	( ) chromosome 9 clo
419	G2155	BZ412041	8.00E-39	Zea mays	OGACG56TC ZM_0.7_1.5_KB Zea mays genomic clone ZMM
419	G2155	gi15528814	3.70E-32	Oryza sativa	hypothetical protein~similar to *Arabidopsis*
419	G2155	gi24059979	1.20E-21	Oryza sativa (japonica cultivar-group)	similar to DNA-bin
419	G2155	gi4165183	3.50E-20	Antirrhinum majus	SAP1 protein.
419	G2155	gi2213534	1.60E-16	Pisum sativum	DNA-binding PD1-like protein.
419	G2155	gi2224911	0.98	Daucus carota	somatic embryogenesis

					receptor-like kinase.
419	G2155	gi454279	1	Avena sativa	DNA-binding protein.
439	G2509	BH989379	8.00E-66	Brassica oleracea	oed22b05.b1 B.oleracea002 Brassica olerac
439	G2509	BQ138607	4.00E-41	Medicago truncatula	NF005C01PH1F1004 Phoma-infected Medicag
439	G2509	BQ786702	4.00E-36	Glycine max	saq72b07.y1 Gm-c1076 Glycine max cDNA clone SOY
439	G2509	OSJN00240	7.00E-31	Oryza sativa	genomic DNA, chromosome 4, BAC clone: OSJNBa0
439	G2509	AAAA01000832	7.00E-31	Oryza sativa (indica cultivar-group)	( ) scaffold000832
439	G2509	BE419451	2.00E-29	Triticum aestivum	WWS012.C2R000101 ITEC WWS Wheat Scutellum
439	G2509	BM062508	5.00E-29	Capsicum annuum	KS01043F09 KS01 Capsicum annuum cDNA, mRNA
439	G2509	AI771755	2.00E-28	Lycopersicon esculentum	EST252855 tomato ovary, TAMU Lycope
439	G2509	CA015575	7.00E-28	Hordeum vulgare subsp. vulgare	HT14L19r HT Hordeum vulgare
439	G2509	BE642320	2.00E-27	Ceratopteris richardii	Cri2_5_L17_SP6 Ceratopteris Spore Li
439	G2509	gi20160854	2.10E-29	Oryza sativa (japonica cultivar-group)	hypothetical prote
439	G2509	gi3264767	8.40E-28	Prunus armeniaca	AP2 domain containing protein.
439	G2509	gi24817250	1.10E-25	Cicer arietinum	transcription factor EREBP-like protein.
439	G2509	gi15217291	7.10E-25	Oryza sativa	Putative AP2 domain containing protein.
439	G2509	gi1208498	1.60E-24	Nicotiana tabacum	EREBP-2.
439	G2509	gi8809571	1.60E-24	Nicotiana sylvestris	ethylene-responsive element binding
439	G2509	gi7528276	3.00E-24	Mesembryanthemum crystallinum	AP2-related transcription f
439	G2509	gi1688233	1.10E-23	Solanum tuberosum	DNA binding protein homolog.
439	G2509	gi4099921	1.60E-23	Stylosanthes hamata	EREBP-3 homolog.
439	G2509	gi18496063	2.40E-23	Fagus sylvatica	ethylene responsive element binding prote
449	G2583	BH658452	1.00E-59	Brassica oleracea	BOMCP74TF BO_2_3_KB Brassica oleracea gen
449	G2583	BE023297	5.00E-54	Glycine max	sm80e10.y1 Gm-c1015 Glycine max cDNA clone GENO
449	G2583	CA486875	1.00E-50	Triticum aestivum	WHE4337_A02_A03ZS Wheat meiotic anther cD
449	G2583	BG642554	8.00E-48	Lycopersicon esculentum	EST356031 tomato flower buds, anthe
449	G2583	BI978981	2.00E-47	Rosa chinensis	zD09 Old Blush petal

					SMART library Rosa chin
449	G2583	BU978490	4.00E-47	Hordeum vulgare subsp. vulgare	HA13G05r HA Hordeum vulgare
449	G2583	BQ106328	4.00E-46	Rosa hybrid cultivar	gg1388.e Rose Petals (Golden Gate) Lam
449	G2583	BI958226	1.00E-44	Hordeum vulgare	HVSMEn0013P17f Hordeum vulgare rachis EST 1
449	G2583	AP004869	1.00E-43	Oryza sativa (japonica cultivar-group)	( ) chromosome 2 clo
449	G2583	BU832200	6.00E-43	Populus tremula x Populus tremuloides	T030G01 Populus apica
449	G2583	gi18650662	2.30E-23	Lycopersicon esculentum	ethylene response factor 1.
449	G2583	gi131754	7.30E-20	Lupinus polyphyllus	PPLZ02 PROTEIN.
449	G2583	gi20160854	2.80E-18	Oryza sativa (japonica cultivar-group)	hypothetical prote
449	G2583	gi10798644	2.80E-18	Nicotiana tabacum	AP2 domain-containing transcription fac
449	G2583	gi8571476	2.80E-18	Atriplex hortensis	apetala2 domain-containing protein.
449	G2583	gi14018047	3.30E-17	Oryza sativa	Putative protein containing AP2 DNA binding
449	G2583	gi12225884	1.10E-16	Zea mays	unnamed protein product.
449	G2583	gi3264767	1.10E-16	Prunus armeniaca	AP2 domain containing protein.
449	G2583	gi4099914	1.10E-16	Stylosanthes hamata	ethylene-responsive element binding p
449	G2583	gi8809573	1.40E-16	Nicotiana sylvestris	ethylene-responsive element binding

Table 12 lists sequences discovered to be paralogous to a number of transcription factors of the present invention. The columns headings include, from left to right, the *Arabidopsis* SEQ ID NO; corresponding *Arabidopsis* Gene ID (GID) numbers; the GID numbers of the paralogs discovered in a database search; and the SEQ ID NOs of the paralogs (a paralog appearing in any cell of the fourth column is also paralogous to the other sequences in that cell).

Table 12. *Arabidopsis* Transcription Factors and Paralogs

SEQ ID NO:	GID NO.	Paralog SEQ ID NO:	Paralog GID No.
10	G28	6, 2074	G6, G1006
12	G47	408	G2133
60	G353	62, 2150, 2156, 2200	G354, G1889, G1974, G2839
88	G481	90, 2010, 2102, 2172	G482, G485, G1364, G2345
94	G489	2054	G714
148	G682	38, 1972, 2142, 2192	G225, G226, G1816, G2718
170	G867	1950, 2072, 370	G9,G993, G1930

186	G912	1958, 1960, 1962, 2162, 2184	G40, G41, G42, G2107, G2513
188	G913	2162	G2107
200	G975	2106, 450	G1387, G2583
224	G1073	2078, 2166	G1067, G2156
226	G1075	2080	G1076
270	G1411	440	G2509
278	G1451	2070	G990
332	G1792	1954, 2134, 2136	G30, G1791, G1795
344	G1836	340	G1818
2158	G1995	64, 66, 2198	G361, G362, G2838
420	G2155	2154	G1945

Table 13 lists the gene identification number (GID) and homologous relationships found using analyses according to Example VIII for the sequences of the Sequence Listing.

Table 13. Homologous relationships found within the Sequence Listing

SEQ ID NO:	GID	DNA or Protein (PRT)	Species from Which Homologous Sequence is Derived	Relationship of SEQ ID NO: to Other Genes
468		DNA	*Glycine max*	Predicted polypeptide sequence is orthologous to G19
469		DNA	*Glycine max*	Predicted polypeptide sequence is orthologous to G19
470		DNA	*Glycine max*	Predicted polypeptide sequence is orthologous to G19
471		DNA	*Glycine max*	Predicted polypeptide sequence is orthologous to G19
472		DNA	*Oryza sativa*	Predicted polypeptide sequence is orthologous to G19
473		DNA	*Oryza sativa*	Predicted polypeptide sequence is orthologous to G19
474		DNA	*Oryza sativa*	Predicted polypeptide sequence is orthologous to G19
475		DNA	*Zea mays*	Predicted polypeptide sequence is orthologous to G19
476		DNA	*Zea mays*	Predicted polypeptide sequence is orthologous to G19
477		DNA	*Glycine max*	Predicted polypeptide sequence is orthologous to G22, G28, G1006
478		DNA	*Glycine max*	Predicted polypeptide sequence is orthologous to G22, G28, G1006
491		DNA	*Glycine max*	Predicted polypeptide sequence is orthologous to G22, G28, G1006
492		DNA	*Glycine max*	Predicted polypeptide sequence is orthologous to G22, G28, G1006
493		DNA	*Glycine max*	Predicted polypeptide sequence is orthologous to G22, G28, G1006
494		DNA	*Glycine max*	Predicted polypeptide sequence is

				orthologous to G22, G28, G1006
495		DNA	*Glycine max*	Predicted polypeptide sequence is orthologous to G22, G28, G1006
496		DNA	*Glycine max*	Predicted polypeptide sequence is orthologous to G22, G28, G1006
497		DNA	*Glycine max*	Predicted polypeptide sequence is orthologous to G22, G28, G1006
498		DNA	*Glycine max*	Predicted polypeptide sequence is orthologous to G22, G28, G1006
499		DNA	*Oryza sativa*	Predicted polypeptide sequence is orthologous to G22, G28, G1006
500		DNA	*Zea mays*	Predicted polypeptide sequence is orthologous to G22, G28, G1006
501		PRT	*Oryza sativa*	Orthologous to G22, G28, G1006
502		PRT	*Oryza sativa*	Orthologous to G22, G28, G1006
503		PRT	*Mesembryanthemum crystallinum*	Orthologous to G22, G28, G1006
504		DNA	*Glycine max*	Predicted polypeptide sequence is orthologous to G47, G2133
505		PRT	*Oryza sativa*	Orthologous to G47, G2133
550		DNA	*Glycine max*	Predicted polypeptide sequence is orthologous to G226, G682, G1816, G2718
551		DNA	*Glycine max*	Predicted polypeptide sequence is orthologous to G226, G682, G1816, G2718
552		DNA	*Glycine max*	Predicted polypeptide sequence is orthologous to G226, G682, G1816, G2718
553		DNA	*Glycine max*	Predicted polypeptide sequence is orthologous to G226, G682, G1816, G2718
554		DNA	*Glycine max*	Predicted polypeptide sequence is orthologous to G226, G682, G1816, G2718
555		DNA	*Oryza sativa*	Predicted polypeptide sequence is orthologous to G226, G682, G1816, G2718
556		DNA	*Zea mays*	Predicted polypeptide sequence is orthologous to G226, G682, G1816, G2718
557		DNA	*Zea mays*	Predicted polypeptide sequence is orthologous to G226, G682, G1816, G2718
558		PRT	*Oryza sativa*	Orthologous to G226, G682, G1816, G2718
559		PRT	*Oryza sativa*	Orthologous to G226, G682, G1816, G2718
610		DNA	*Glycine max*	Predicted polypeptide sequence is orthologous to G353, G354, G1974, G1889, G2839
611		DNA	*Glycine max*	Predicted polypeptide sequence is orthologous to G353, G354, G1974, G1889, G2839
612		DNA	*Glycine max*	Predicted polypeptide sequence is orthologous to G353, G354, G1974, G1889, G2839
613		DNA	*Oryza sativa*	Predicted polypeptide sequence is orthologous to G353, G354, G1974, G1889, G2839
614		DNA	*Zea mays*	Predicted polypeptide sequence is orthologous to G353, G354, G1974, G1889, G2839

615		DNA	*Zea mays*	Predicted polypeptide sequence is orthologous to G353, G354, G1974, G1889, G2839
616		DNA	*Zea mays*	Predicted polypeptide sequence is orthologous to G353, G354, G1974, G1889, G2839
617		DNA	*Zea mays*	Predicted polypeptide sequence is orthologous to G353, G354, G1974, G1889, G2839
618		DNA	*Zea mays*	Predicted polypeptide sequence is orthologous to G353, G354, G1974, G1889, G2839
619		DNA	*Zea mays*	Predicted polypeptide sequence is orthologous to G353, G354, G1974, G1889, G2839
620		DNA	*Zea mays*	Predicted polypeptide sequence is orthologous to G353, G354, G1974, G1889, G2839
621		PRT	*Oryza sativa*	Orthologous to G353, G354, G1974, G1889, G2839
622		PRT	*Oryza sativa*	Orthologous to G353, G354, G1974, G1889, G2839
623		PRT	*Oryza sativa*	Orthologous to G353, G354, G1974, G1889, G2839
624		PRT	*Oryza sativa*	Orthologous to G353, G354, G1974, G1889, G2839
625		PRT	*Oryza sativa*	Orthologous to G353, G354, G1974, G1889, G2839
626		PRT	*Oryza sativa*	Orthologous to G353, G354, G1974, G1889, G2839
746		DNA	*Glycine max*	Predicted polypeptide sequence is orthologous to G481, G482, G485, G1364, G2345
747		DNA	*Glycine max*	Predicted polypeptide sequence is orthologous to G481, G482, G485, G1364, G2345
748		DNA	*Glycine max*	Predicted polypeptide sequence is orthologous to G481, G482, G485, G1364, G2345
749		DNA	*Glycine max*	Predicted polypeptide sequence is orthologous to G481, G482, G485, G1364, G2345
750		DNA	*Glycine max*	Predicted polypeptide sequence is orthologous to G481, G482, G485, G1364, G2345
751		DNA	*Glycine max*	Predicted polypeptide sequence is orthologous to G481, G482, G485, G1364, G2345
752		DNA	*Glycine max*	Predicted polypeptide sequence is orthologous to G481, G482, G485, G1364, G2345
753		DNA	*Glycine max*	Predicted polypeptide sequence is orthologous to G481, G482, G485, G1364, G2345

754		DNA	*Glycine max*	Predicted polypeptide sequence is orthologous to G481, G482, G485, G1364, G2345
755		DNA	*Glycine max*	Predicted polypeptide sequence is orthologous to G481, G482, G485, G1364, G2345
756		DNA	*Oryza sativa*	Predicted polypeptide sequence is orthologous to G481, G482, G485, G1364, G2345
757		DNA	*Oryza sativa*	Predicted polypeptide sequence is orthologous to G481, G482, G485, G1364, G2345
758		DNA	*Zea mays*	Predicted polypeptide sequence is orthologous to G481, G482, G485, G1364, G2345
759		DNA	*Zea mays*	Predicted polypeptide sequence is orthologous to G481, G482, G485, G1364, G2345
760		DNA	*Zea mays*	Predicted polypeptide sequence is orthologous to G481, G482, G485, G1364, G2345
761		DNA	*Zea mays*	Predicted polypeptide sequence is orthologous to G481, G482, G485, G1364, G2345
762		DNA	*Zea mays*	Predicted polypeptide sequence is orthologous to G481, G482, G485, G1364, G2345
763		DNA	*Zea mays*	Predicted polypeptide sequence is orthologous to G481, G482, G485, G1364, G2345
764		DNA	*Zea mays*	Predicted polypeptide sequence is orthologous to G481, G482, G485, G1364, G2345
765		DNA	*Zea mays*	Predicted polypeptide sequence is orthologous to G481, G482, G485, G1364, G2345
766		DNA	*Zea mays*	Predicted polypeptide sequence is orthologous to G481, G482, G485, G1364, G2345
767		DNA	*Zea mays*	Predicted polypeptide sequence is orthologous to G481, G482, G485, G1364, G2345
768		DNA	*Gossypium arboreum*	Predicted polypeptide sequence is orthologous to G481, G482, G485, G1364, G2345
769		DNA	*Glycine max*	Predicted polypeptide sequence is orthologous to G481, G482, G485, G1364, G2345
770		DNA	*Gossypium hirsutum*	Predicted polypeptide sequence is orthologous to G481, G482, G485, G1364, G2345
771		DNA	*Lycopersicon esculentum*	Predicted polypeptide sequence is orthologous to G481, G482, G485, G1364, G2345

772		DNA	*Lycopersicon esculentum*	Predicted polypeptide sequence is orthologous to G481, G482, G485, G1364, G2345
773		DNA	*Medicago truncatula*	Predicted polypeptide sequence is orthologous to G481, G482, G485, G1364, G2345
774		DNA	*Lycopersicon esculentum*	Predicted polypeptide sequence is orthologous to G481, G482, G485, G1364, G2345
775		DNA	*Solanum tuberosum*	Predicted polypeptide sequence is orthologous to G481, G482, G485, G1364, G2345
776		DNA	*Triticum aestivum*	Predicted polypeptide sequence is orthologous to G481, G482, G485, G1364, G2345
777		DNA	*Hordeum vulgare*	Predicted polypeptide sequence is orthologous to G481, G482, G485, G1364, G2345
778		DNA	*Triticum monococcum*	Predicted polypeptide sequence is orthologous to G481, G482, G485, G1364, G2345
779		DNA	*Glycine max*	Predicted polypeptide sequence is orthologous to G481, G482, G485, G1364, G2345
780		PRT	*Oryza sativa*	Orthologous to G481, G482, G485, G1364, G2345
781		PRT	*Oryza sativa*	Orthologous to G481, G482, G485, G1364, G2345
782		PRT	*Oryza sativa*	Orthologous to G481, G482, G485, G1364, G2345
783		PRT	*Oryza sativa*	Orthologous to G481, G482, G485, G1364, G2345
784		PRT	*Oryza sativa*	Orthologous to G481, G482, G485, G1364, G2345
785		PRT	*Zea mays*	Orthologous to G481, G482, G485, G1364, G2345
786		PRT	*Zea mays*	Orthologous to G481, G482, G485, G1364, G2345
787		PRT	*Oryza sativa*	Orthologous to G481, G482, G485, G1364, G2345
788		PRT	*Oryza sativa*	Orthologous to G481, G482, G485, G1364, G2345
789		PRT	*Oryza sativa*	Orthologous to G481, G482, G485, G1364, G2345
790		PRT	*Oryza sativa*	Orthologous to G481, G482, G485, G1364, G2345
791		PRT	*Oryza sativa*	Orthologous to G481, G482, G485, G1364, G2345
792		PRT	*Oryza sativa*	Orthologous to G481, G482, G485, G1364, G2345
793		PRT	*Oryza sativa*	Orthologous to G481, G482, G485, G1364, G2345
794		PRT	*Oryza sativa*	Orthologous to G481, G482, G485, G1364, G2345

795		PRT	*Oryza sativa*	Orthologous to G481, G482, G485, G1364, G2345
796		PRT	*Oryza sativa*	Orthologous to G481, G482, G485, G1364, G2345
797		PRT	*Glycine max*	Orthologous to G481, G482, G485, G1364, G2345
798		PRT	*Glycine max*	Orthologous to G481, G482, G485, G1364, G2345
799		PRT	*Glycine max*	Orthologous to G481, G482, G485, G1364, G2345
800		PRT	*Glycine max*	Orthologous to G481, G482, G485, G1364, G2345
801		PRT	*Glycine max*	Orthologous to G481, G482, G485, G1364, G2345
802		PRT	*Glycine max*	Orthologous to G481, G482, G485, G1364, G2345
803		PRT	*Glycine max*	Orthologous to G481, G482, G485, G1364, G2345
804		PRT	*Zea mays*	Orthologous to G481, G482, G485, G1364, G2345
805		PRT	*Zea mays*	Orthologous to G481, G482, G485, G1364, G2345
806		PRT	*Zea mays*	Orthologous to G481, G482, G485, G1364, G2345
807		PRT	*Zea mays*	Orthologous to G481, G482, G485, G1364, G2345
825		DNA	*Glycine max*	Predicted polypeptide sequence is orthologous to G489, G714
826		DNA	*Glycine max*	Predicted polypeptide sequence is orthologous to G489, G714
827		DNA	*Glycine max*	Predicted polypeptide sequence is orthologous to G489, G714
828		DNA	*Glycine max*	Predicted polypeptide sequence is orthologous to G489, G714
829		DNA	*Glycine max*	Predicted polypeptide sequence is orthologous to G489, G714
830		DNA	*Glycine max*	Predicted polypeptide sequence is orthologous to G489, G714
831		DNA	*Glycine max*	Predicted polypeptide sequence is orthologous to G489, G714
832		DNA	*Oryza sativa*	Predicted polypeptide sequence is orthologous to G489, G714
833		DNA	*Oryza sativa*	Predicted polypeptide sequence is orthologous to G489, G714
834		DNA	*Zea mays*	Predicted polypeptide sequence is orthologous to G489, G714
835		PRT	*Oryza sativa*	Orthologous to G489, G714
836		PRT	*Oryza sativa*	Orthologous to G489, G714
837		PRT	*Oryza sativa*	Orthologous to G489, G714
981		DNA	*Oryza sativa*	Predicted polypeptide sequence is orthologous to G634
982		DNA	*Oryza sativa*	Predicted polypeptide sequence is orthologous to G634

983		DNA	*Oryza sativa*	Predicted polypeptide sequence is orthologous to G634
984		DNA	*Zea mays*	Predicted polypeptide sequence is orthologous to G634
985		DNA	*Zea mays*	Predicted polypeptide sequence is orthologous to G634
986		DNA	*Zea mays*	Predicted polypeptide sequence is orthologous to G634
987		PRT	*Oryza sativa*	Orthologous to G634
988		PRT	*Oryza sativa*	Orthologous to G634
1076		DNA	*Glycine max*	Predicted polypeptide sequence is orthologous to G226, G682, G1816, G2718
1077		DNA	*Hordeum vulgare subsp. vulgare*	Predicted polypeptide sequence is orthologous to G226, G682, G1816, G2718
1078		DNA	*Populus tremula x Populus tremuloides*	Predicted polypeptide sequence is orthologous to G226, G682, G1816, G2718
1079		DNA	*Triticum aestivum*	Predicted polypeptide sequence is orthologous to G226, G682, G1816, G2718
1080		DNA	*Gossypium arboreum*	Predicted polypeptide sequence is orthologous to G226, G682, G1816, G2718
1081		PRT	*Oryza sativa*	Orthologous to G226, G682, G1816, G2718
1082		PRT	*Oryza sativa*	Orthologous to G226, G682, G1816, G2718
1083		PRT	*Glycine max*	Orthologous to G226, G682, G1816, G2718
1084		PRT	*Glycine max*	Orthologous to G226, G682, G1816, G2718
1085		PRT	*Glycine max*	Orthologous to G226, G682, G1816, G2718
1086		PRT	*Glycine max*	Orthologous to G226, G682, G1816, G2718
1087		PRT	*Glycine max*	Orthologous to G226, G682, G1816, G2718
1088		PRT	*Glycine max*	Orthologous to G226, G682, G1816, G2718
1089		PRT	*Zea mays*	Orthologous to G226, G682, G1816, G2718
1090		PRT	*Zea mays*	Orthologous to G226, G682, G1816, G2718
1159		DNA	*Glycine max*	Predicted polypeptide sequence is orthologous to G9, G867, G993, G1930
1160		DNA	*Glycine max*	Predicted polypeptide sequence is orthologous to G9, G867, G993, G1930
1161		DNA	*Glycine max*	Predicted polypeptide sequence is orthologous to G9, G867, G993, G1930
1162		DNA	*Glycine max*	Predicted polypeptide sequence is orthologous to G9, G867, G993, G1930
1163		DNA	*Glycine max*	Predicted polypeptide sequence is orthologous to G9, G867, G993, G1930
1164		DNA	*Glycine max*	Predicted polypeptide sequence is orthologous to G9, G867, G993, G1930
1165		DNA	*Oryza sativa*	Predicted polypeptide sequence is orthologous to G9, G867, G993, G1930
1166		DNA	*Oryza sativa*	Predicted polypeptide sequence is orthologous to G9, G867, G993, G1930
1167		DNA	*Zea mays*	Predicted polypeptide sequence is orthologous to G9, G867, G993, G1930
1168		DNA	*Zea mays*	Predicted polypeptide sequence is orthologous to G9, G867, G993, G1930
1169		DNA	*Zea mays*	Predicted polypeptide sequence is orthologous to G9, G867, G993, G1930
1170		DNA	*Zea mays*	Predicted polypeptide sequence is

				orthologous to G9, G867, G993, G1930
1171		DNA	*Glycine max*	Predicted polypeptide sequence is orthologous to G9, G867, G993, G1930
1172		DNA	*Mesembryanthemum crystallinum*	Predicted polypeptide sequence is orthologous to G9, G867, G993, G1930
1173		DNA	*Lycopersicon esculentum*	Predicted polypeptide sequence is orthologous to G9, G867, G993, G1930
1174		DNA	*Solanum tuberosum*	Predicted polypeptide sequence is orthologous to G9, G867, G993, G1930
1175		DNA	*Hordeum vulgare*	Predicted polypeptide sequence is orthologous to G9, G867, G993, G1930
1176		PRT	*Oryza sativa*	Orthologous to G9, G867, G993, G1930
1177		PRT	*Oryza sativa*	Orthologous to G9, G867, G993, G1930
1178		PRT	*Oryza sativa*	Orthologous to G9, G867, G993, G1930
1179		PRT	*Oryza sativa*	Orthologous to G9, G867, G993, G1930
1180		PRT	*Oryza sativa*	Orthologous to G9, G867, G993, G1930
1181		PRT	*Oryza sativa*	Orthologous to G9, G867, G993, G1930
1182		PRT	*Glycine max*	Orthologous to G9, G867, G993, G1930
1183		PRT	*Glycine max*	Orthologous to G9, G867, G993, G1930
1184		PRT	*Glycine max*	Orthologous to G9, G867, G993, G1930
1185		PRT	*Zea mays*	Orthologous to G9, G867, G993, G1930
1186		PRT	*Zea mays*	Orthologous to G9, G867, G993, G1930
1204		DNA	*Glycine max*	Predicted polypeptide sequence is orthologous to G912, G2107, G2513
1205		DNA	*Glycine max*	Predicted polypeptide sequence is orthologous to G912, G2107, G2513
1206		DNA	*Glycine max*	Predicted polypeptide sequence is orthologous to G912, G2107, G2513
1207		DNA	*Glycine max*	Predicted polypeptide sequence is orthologous to G912, G2107, G2513
1208		DNA	*Glycine max*	Predicted polypeptide sequence is orthologous to G912, G2107, G2513
1209		DNA	*Glycine max*	Predicted polypeptide sequence is orthologous to G912, G2107, G2513
1210		DNA	*Glycine max*	Predicted polypeptide sequence is orthologous to G912, G2107, G2513
1211		DNA	*Oryza sativa*	Predicted polypeptide sequence is orthologous to G912, G2107, G2513
1212		DNA	*Oryza sativa*	Predicted polypeptide sequence is orthologous to G912, G913
1213		DNA	*Zea mays*	Predicted polypeptide sequence is orthologous to G912, G2107, G2513
1214		DNA	*Zea mays*	Predicted polypeptide sequence is orthologous to G912, G2107, G2513
1215		DNA	*Zea mays*	Predicted polypeptide sequence is orthologous to G912, G913
1216		DNA	*Zea mays*	Predicted polypeptide sequence is orthologous to G912, G2107, G2513
1217		DNA	*Zea mays*	Predicted polypeptide sequence is orthologous to G912, G2107, G2513
1218		DNA	*Brassica napus*	Predicted polypeptide sequence is orthologous to G912, G913
1219		DNA	*Solanum tuberosum*	Predicted polypeptide sequence is

				orthologous to G912, G2107, G2513
1220		DNA	*Descurainia sophia*	Predicted polypeptide sequence is orthologous to G912, G2107, G2513
1221		PRT	*Oryza sativa*	Orthologous to G912, G2107, G2513
1222		PRT	*Oryza sativa*	Orthologous to G912, G913
1223		PRT	*Oryza sativa*	Orthologous to G912, G913
1224		PRT	*Oryza sativa*	Orthologous to G912, G2107, G2513
1225		PRT	*Brassica napus*	Orthologous to G912, G2107, G2513
1226		PRT	*Nicotiana tabacum*	Orthologous to G912, G2107, G2513
1227		PRT	*Oryza sativa*	Orthologous to G912, G2107, G2513
1228		PRT	*Oryza sativa*	Orthologous to G912, G2107, G2513
1229		PRT	*Oryza sativa*	Orthologous to G912, G2107, G2513
1230		PRT	*Oryza sativa*	Orthologous to G912, G2107, G2513
1231		PRT	*Oryza sativa*	Orthologous to G912, G2107, G2513
1232		PRT	*Oryza sativa*	Orthologous to G912, G2107, G2513
1233		PRT	*Oryza sativa*	Orthologous to G912, G2107, G2513
1234		PRT	*Oryza sativa*	Orthologous to G912, G2107, G2513
1235		PRT	*Oryza sativa*	Orthologous to G912, G2107, G2513
1236		PRT	*Oryza sativa*	Orthologous to G912, G2107, G2513
1237		PRT	*Glycine max*	Orthologous to G912, G2107, G2513
1238		PRT	*Glycine max*	Orthologous to G912, G2107, G2513
1239		PRT	*Glycine max*	Orthologous to G912, G2107, G2513
1240		PRT	*Glycine max*	Orthologous to G912, G2107, G2513
1241		PRT	*Glycine max*	Orthologous to G912, G2107, G2513
1242		PRT	*Glycine max*	Orthologous to G912, G2107, G2513
1243		PRT	*Glycine max*	Orthologous to G912, G2107, G2513
1244		PRT	*Zea mays*	Orthologous to G912, G2107, G2513
1245		PRT	*Zea mays*	Orthologous to G912, G2107, G2513
1246		PRT	*Zea mays*	Orthologous to G912, G2107, G2513
1247		PRT	*Zea mays*	Orthologous to G912, G2107, G2513
1248		PRT	*Zea mays*	Orthologous to G912, G2107, G2513
1249		DNA	*Glycine max*	Predicted polypeptide sequence is orthologous to G922
1250		DNA	*Glycine max*	Predicted polypeptide sequence is orthologous to G922
1251		DNA	*Glycine max*	Predicted polypeptide sequence is orthologous to G922
1252		DNA	*Oryza sativa*	Predicted polypeptide sequence is orthologous to G922
1253		DNA	*Oryza sativa*	Predicted polypeptide sequence is orthologous to G922
1254		PRT	*Oryza sativa*	Orthologous to G922
1255		PRT	*Oryza sativa*	Orthologous to G922
1256		PRT	*Oryza sativa*	Orthologous to G922
1257		PRT	*Oryza sativa*	Orthologous to G922
1258		DNA	*Glycine max*	Predicted polypeptide sequence is orthologous to G926
1259		DNA	*Glycine max*	Predicted polypeptide sequence is orthologous to G926
1260		DNA	*Oryza sativa*	Predicted polypeptide sequence is orthologous to G926

1261		DNA	*Oryza sativa*	Predicted polypeptide sequence is orthologous to G926
1262		DNA	*Zea mays*	Predicted polypeptide sequence is orthologous to G926
1263		PRT	*Brassica napus*	Orthologous to G926
1292		DNA	*Glycine max*	Predicted polypeptide sequence is orthologous to G975, G1387, G2583
1293		DNA	*Glycine max*	Predicted polypeptide sequence is orthologous to G975, G1387, G2583
1294		DNA	*Glycine max*	Predicted polypeptide sequence is orthologous to G975, G1387, G2583
1295		DNA	*Glycine max*	Predicted polypeptide sequence is orthologous to G975, G1387, G2583
1296		DNA	*Glycine max*	Predicted polypeptide sequence is orthologous to G975, G1387, G2583
1297		DNA	*Oryza sativa*	Predicted polypeptide sequence is orthologous to G975, G1387, G2583
1298		DNA	*Oryza sativa*	Predicted polypeptide sequence is orthologous to G975, G1387, G2583
1299		DNA	*Zea mays*	Predicted polypeptide sequence is orthologous to G975, G1387, G2583
1300		DNA	*Zea mays*	Predicted polypeptide sequence is orthologous to G975, G1387, G2583
1301		DNA	*Brassica rapa*	Predicted polypeptide sequence is orthologous to G975, G1387, G2583
1302		PRT	*Oryza sativa*	Orthologous to G975, G1387, G2583
1393		DNA	*Glycine max*	Predicted polypeptide sequence is orthologous to G1069, G2153
1394		DNA	*Glycine max*	Predicted polypeptide sequence is orthologous to G1069, G2153
1395		PRT	*Oryza sativa*	Orthologous to G1069, G2153
1396		DNA	*Zea mays*	Predicted polypeptide sequence is orthologous to G1069, G2153
1397		DNA	*Lotus japonicus*	Predicted polypeptide sequence is orthologous to G1069, G2153
1398		DNA	*Lycopersicon esculentum*	Predicted polypeptide sequence is orthologous to G1067, G1073, G2156
1399		PRT	*Oryza sativa*	Orthologous to G1067, G1073, G2156
1400		PRT	*Oryza sativa*	Orthologous to G1067, G1073, G2156
1401		PRT	*Oryza sativa*	Orthologous to G1067, G1073, G2156
1402		PRT	*Oryza sativa*	Orthologous to G1067, G1073, G2156
1403		PRT	*Oryza sativa*	Orthologous to G1067, G1073, G2156
1404		PRT	*Oryza sativa*	Orthologous to G1067, G1073, G2156
1405		PRT	*Oryza sativa*	Orthologous to G1067, G1073, G2156
1406		PRT	*Oryza sativa*	Orthologous to G1067, G1073, G2156
1407		PRT	*Oryza sativa*	Orthologous to G1067, G1073, G2156
1408		PRT	*Oryza sativa*	Orthologous to G1067, G1073, G2156
1409		PRT	*Oryza sativa*	Orthologous to G1067, G1073, G2156
1410		PRT	*Oryza sativa*	Orthologous to G1067, G1073, G2156
1411		PRT	*Glycine max*	Orthologous to G1067, G1073, G2156
1412		PRT	*Glycine max*	Orthologous to G1067, G1073, G2156
1413		PRT	*Glycine max*	Orthologous to G1067, G1073, G2156

1414		PRT	*Glycine max*	Orthologous to G1067, G1073, G2156
1415		PRT	*Glycine max*	Orthologous to G1067, G1073, G2156
1416		PRT	*Glycine max*	Orthologous to G1067, G1073, G2156
1417		PRT	*Glycine max*	Orthologous to G1067, G1073, G2156
1418		PRT	*Zea mays*	Orthologous to G1067, G1073, G2156
1419		DNA	*Glycine max*	Predicted polypeptide sequence is orthologous to G1075, G1076
1420		DNA	*Glycine max*	Predicted polypeptide sequence is orthologous to G1075, G1076
1421		DNA	*Glycine max*	Predicted polypeptide sequence is orthologous to G1075, G1076
1422		DNA	*Glycine max*	Predicted polypeptide sequence is orthologous to G1075, G1076
1423		DNA	*Glycine max*	Predicted polypeptide sequence is orthologous to G1075, G1076
1424		DNA	*Oryza sativa*	Predicted polypeptide sequence is orthologous to G1075, G1076
1425		DNA	*Oryza sativa*	Predicted polypeptide sequence is orthologous to G1075, G1076
1426		DNA	*Oryza sativa*	Predicted polypeptide sequence is orthologous to G1075, G1076
1587		DNA	*Glycine max*	Predicted polypeptide sequence is orthologous to G1411, G2509
1588		DNA	*Glycine max*	Predicted polypeptide sequence is orthologous to G1411, G2509
1589		DNA	*Glycine max*	Predicted polypeptide sequence is orthologous to G1411, G2509
1590		DNA	*Glycine max*	Predicted polypeptide sequence is orthologous to G1411, G2509
1591		DNA	*Zea mays*	Predicted polypeptide sequence is orthologous to G1411, G2509
1604		DNA	*Glycine max*	Predicted polypeptide sequence is orthologous to G990, G1451
1605		DNA	*Glycine max*	Predicted polypeptide sequence is orthologous to G990, G1451
1606		DNA	*Oryza sativa*	Predicted polypeptide sequence is orthologous to G990, G1451
1607		DNA	*Oryza sativa*	Predicted polypeptide sequence is orthologous to G990, G1451
1608		DNA	*Oryza sativa*	Predicted polypeptide sequence is orthologous to G990, G1451
1609		DNA	*Zea mays*	Predicted polypeptide sequence is orthologous to G990, G1451
1610		DNA	*Zea mays*	Predicted polypeptide sequence is orthologous to G990, G1451
1611		DNA	*Zea mays*	Predicted polypeptide sequence is orthologous to G990, G1451
1612		DNA	*Zea mays*	Predicted polypeptide sequence is orthologous to G990, G1451
1613		DNA	*Medicago truncatula*	Predicted polypeptide sequence is orthologous to G990, G1451
1614		DNA	*Solanum tuberosum*	Predicted polypeptide sequence is orthologous to G990, G1451

1615		DNA	*Zea mays*	Predicted polypeptide sequence is orthologous to G990, G1451
1616		DNA	*Sorghum propinquum*	Predicted polypeptide sequence is orthologous to G990, G1451
1617		DNA	*Glycine max*	Predicted polypeptide sequence is orthologous to G990, G1451
1618		DNA	*Sorghum bicolor*	Predicted polypeptide sequence is orthologous to G990, G1451
1619		DNA	*Hordeum vulgare*	Predicted polypeptide sequence is orthologous to G990, G1451
1620		DNA	*Lycopersicon esculentum*	Predicted polypeptide sequence is orthologous to G990, G1451
1621		PRT	*Oryza sativa*	Orthologous to G990, G1451
1622		PRT	*Oryza sativa*	Orthologous to G990, G1451
1623		PRT	*Oryza sativa*	Orthologous to G990, G1451
1624		PRT	*Oryza sativa*	Orthologous to G990, G1451
1671		DNA	*Glycine max*	Predicted polypeptide sequence is orthologous to G1543
1672		DNA	*Oryza sativa*	Predicted polypeptide sequence is orthologous to G1543
1673		DNA	*Zea mays*	Predicted polypeptide sequence is orthologous to G1543
1674		PRT	*Oryza sativa*	Orthologous to G1543
1728		DNA	*Glycine max*	Predicted polypeptide sequence is orthologous to G30, G1791, G1792, G1795
1729		DNA	*Glycine max*	Predicted polypeptide sequence is orthologous to G30, G1791, G1792, G1795
1730		DNA	*Glycine max*	Predicted polypeptide sequence is orthologous to G30, G1791, G1792, G1795
1731		DNA	*Glycine max*	Predicted polypeptide sequence is orthologous to G30, G1791, G1792, G1795
1732		DNA	*Glycine max*	Predicted polypeptide sequence is orthologous to G30, G1791, G1792, G1795
1733		DNA	*Zea mays*	Predicted polypeptide sequence is orthologous to G30, G1791, G1792, G1795
1734		DNA	*Lycopersicon esculentum*	Predicted polypeptide sequence is orthologous to G30, G1791, G1792, G1795
1735	G3380	PRT	*Oryza sativa*	Orthologous to G30, G1791, G1792, G1795
1736	G3381	PRT	*Oryza sativa indica*	Orthologous to G30, G1791, G1792, G1795
1737	G3383	PRT	*Oryza sativa japonica*	Orthologous to G30, G1791, G1792, G1795
1795		DNA	*Oryza sativa*	Predicted polypeptide sequence is orthologous to G9, G867, G993, G1930
1908		DNA	*Medicago truncatula*	Predicted polypeptide sequence is orthologous to G1945, G2155
1909		DNA	*Medicago truncatula*	Predicted polypeptide sequence is orthologous to G1945, G2155
1910		DNA	*Glycine max*	Predicted polypeptide sequence is orthologous to G1945, G2155
1949	G9	DNA	*Arabidopsis thaliana*	Predicted polypeptide sequence is paralogous to G9, G867, G993, G1930
1950	G9	PRT	*Arabidopsis thaliana*	Paralogous to G9, G867, G993, G1930
1953	G30	DNA	*Arabidopsis thaliana*	Predicted polypeptide sequence is paralogous to G30, G1791, G1792, G1795

1954	G30	PRT	*Arabidopsis thaliana*	start Paralogous to G30, G1791, G1792, G1795
1955	G40	DNA	*Arabidopsis thaliana*	Predicted polypeptide sequence is paralogous to G912, G2107, G2513
1956	G40	PRT	*Arabidopsis thaliana*	Paralogous to G912, G2107, G2513
1957	G41	DNA	*Arabidopsis thaliana*	Predicted polypeptide sequence is paralogous to G912, G2107, G2513
1958	G41	PRT	*Arabidopsis thaliana*	Paralogous to G912, G2107, G2513
1959	G42	DNA	*Arabidopsis thaliana*	Predicted polypeptide sequence is paralogous to G912, G2107, G2513
1960	G42	PRT	*Arabidopsis thaliana*	Paralogous to G912, G2107, G2513
1971	G225	DNA	*Arabidopsis thaliana*	Predicted polypeptide sequence is paralogous to G226, G682, G1816, G2718
1972	G225	PRT	*Arabidopsis thaliana*	Paralogous to G226, G682, G1816, G2718
1991	G370	DNA	*Arabidopsis thaliana*	Predicted polypeptide sequence is paralogous to G361, G362, G370, G1995, G2826, G2838
1992	G370	PRT	*Arabidopsis thaliana*	Paralogous to G361, G362, G370, G1995, G2826, G2838
2009	G485	DNA	*Arabidopsis thaliana*	Predicted polypeptide sequence is paralogous to 481, G482, G485, G1364, G2345
2010	G485	PRT	*Arabidopsis thaliana*	Paralogous to 481, G482, G485, G1364, G2345
2053	G714	DNA	*Arabidopsis thaliana*	Predicted polypeptide sequence is paralogous to G489, G714
2054	G714	PRT	*Arabidopsis thaliana*	Paralogous to G489, G714
2069	G990	DNA	*Arabidopsis thaliana*	Predicted polypeptide sequence is paralogous to G990, G1451
2070	G990	PRT	*Arabidopsis thaliana*	Paralogous to G990, G1451
2071	G993	DNA	*Arabidopsis thaliana*	Predicted polypeptide sequence is paralogous to G9, G867, G993, G1930
2072	G993	PRT	*Arabidopsis thaliana*	Paralogous to G9, G867, G993, G1930
2073	G1006	DNA	*Arabidopsis thaliana*	Predicted polypeptide sequence is paralogous to G22, G28, G1006
2074	G1006	PRT	*Arabidopsis thaliana*	Paralogous to G22, G28, G1006
2077	G1067	DNA	*Arabidopsis thaliana*	Predicted polypeptide sequence is paralogous to G1067, G1073, G2156
2078	G1067	PRT	*Arabidopsis thaliana*	Paralogous to G1067, G1073, G2156
2079	G1076	DNA	*Arabidopsis thaliana*	Predicted polypeptide sequence is paralogous to G1075, G1076
2080	G1076	PRT	*Arabidopsis thaliana*	Paralogous to G1075, G1076
2101	G1364	DNA	*Arabidopsis thaliana*	Predicted polypeptide sequence is paralogous to 481, G482, G485, G1364, G2345
2102	G1364	PRT	*Arabidopsis thaliana*	Paralogous to 481, G482, G485, G1364, G2345
2105	G1387	DNA	*Arabidopsis thaliana*	Predicted polypeptide sequence is paralogous to G975, G1387, G2583
2106	G1387	PRT	*Arabidopsis thaliana*	Paralogous to G975, G1387, G2583
2133	G1791	DNA	*Arabidopsis thaliana*	Predicted polypeptide sequence is paralogous to G30, G1791, G1792, G1795
2134	G1791	PRT	*Arabidopsis thaliana*	Paralogous to G30, G1791, G1792, G1795

2135	G1795	DNA	*Arabidopsis thaliana*	Predicted polypeptide sequence is paralogous to G30, G1791, G1792, G1795
2136	G1795	PRT	*Arabidopsis thaliana*	Paralogous to G30, G1791, G1792, G1795
2141	G1816	DNA	*Arabidopsis thaliana*	Predicted polypeptide sequence is paralogous to G226, G682, G1816, G2718
2142	G1816	PRT	*Arabidopsis thaliana*	Paralogous to G226, G682, G1816, G2718
2149	G1889	DNA	*Arabidopsis thaliana*	Predicted polypeptide sequence is paralogous to G353, G354, G1974, G1889, G2839
2150	G1889	PRT	*Arabidopsis thaliana*	Paralogous to G353, G354, G1974, G1889, G2839
2153	G1945	DNA	*Arabidopsis thaliana*	Predicted polypeptide sequence is paralogous to G1945, G2155
2154	G1945	PRT	*Arabidopsis thaliana*	Paralogous to G1945, G2155
2155	G1974	DNA	*Arabidopsis thaliana*	Predicted polypeptide sequence is paralogous to G353, G354, G1974, G1889, G2839
2156	G1974	PRT	*Arabidopsis thaliana*	Paralogous to G353, G354, G1974, G1889, G2839
2157	G1995	DNA	*Arabidopsis thaliana*	Predicted polypeptide sequence is paralogous to G361, G362, G1995, G2826, G2838
2158	G1995	PRT	*Arabidopsis thaliana*	Paralogous to G361, G362, G1995, G2826, G2838
2161	G2107	DNA	*Arabidopsis thaliana*	Predicted polypeptide sequence is paralogous to G912, G2107, G2513
2162	G2107	PRT	*Arabidopsis thaliana*	Paralogous to G912, G2107, G2513
2165	G2156	DNA	*Arabidopsis thaliana*	Predicted polypeptide sequence is paralogous to G1067, G1073, G2156
2166	G2156	PRT	*Arabidopsis thaliana*	Paralogous to G1067, G1073, G2156
2171	G2345	DNA	*Arabidopsis thaliana*	Predicted polypeptide sequence is paralogous to 481, G482, G485, G1364, G2345
2172	G2345	PRT	*Arabidopsis thaliana*	Paralogous to 481, G482, G485, G1364, G2345
2183	G2513	DNA	*Arabidopsis thaliana*	Predicted polypeptide sequence is paralogous to G912, G2107, G2513
2184	G2513	PRT	*Arabidopsis thaliana*	Paralogous to G912, G2107, G2513
2191	G2718	DNA	*Arabidopsis thaliana*	Predicted polypeptide sequence is paralogous to G226, G682, G1816, G2718
2192	G2718	PRT	*Arabidopsis thaliana*	Paralogous to G226, G682, G1816, G2718
2199	G2839	DNA	*Arabidopsis thaliana*	Predicted polypeptide sequence is paralogous to G353, G354, G1974, G1889, G2839
2200	G2839	PRT	*Arabidopsis thaliana*	Paralogous to G353, G354, G1974, G1889, G2839

Molecular Modeling

Another means that may be used to confirm the utility and function of transcription factor sequences that are orthologous or paralogous to presently disclosed transcription factors is through the use of molecular modeling software. Molecular modeling is routinely used to predict polypeptide

structure, and a variety of protein structure modeling programs, such as "Insight II" (Accelrys, Inc.) are commercially available for this purpose. Modeling can thus be used to predict which residues of a polypeptide can be changed without altering function (Crameri et al. (2003) U.S. Patent No. 6, 521, 453). Thus, polypeptides that are sequentially similar can be shown to have a high likelihood of similar function by their structural similarity, which may, for example, be established by comparison of regions of superstructure. The relative tendencies of amino acids to form regions of superstructure (for example, helixes and β-sheets) are well established. For example, O'Neil et al. (1990) *Science* 250: 646-651) have discussed in detail the helix forming tendencies of amino acids. Tables of relative structure forming activity for amino acids can be used as substitution tables to predict which residues can be functionally substituted in a given region, for example, in DNA-binding domains of known transcription factors and equivalogs. Homologs that are likely to be functionally similar can then be identified.

Of particular interest is the structure of a transcription factor in the region of its conserved domain, such as those identified in Table 8. Structural analyses may be performed by comparing the structure of the known transcription factor around its conserved domain with those of orthologs and paralogs. Analysis of a number of polypeptides within a transcription factor group or clade, including the functionally or sequentially similar polypeptides provided in the Sequence Listing, may also provide an understanding of structural elements required to regulate transcription within a given family.

EXAMPLES

The invention, now being generally described, will be more readily understood by reference to the following examples, which are included merely for purposes of illustration of certain aspects and embodiments of the present invention and are not intended to limit the invention. It will be recognized by one of skill in the art that a transcription factor that is associated with a particular first trait may also be associated with at least one other, unrelated and inherent second trait which was not predicted by the first trait.

The complete descriptions of the traits associated with each polynucleotide of the invention are fully disclosed in Table 7 and Table 9. The complete description of the transcription factor gene family and identified conserved domains of the polypeptide encoded by the polynucleotide is fully disclosed in Table 8.

Example I: Full Length Gene Identification and Cloning

Putative transcription factor sequences (genomic or ESTs) related to known transcription factors were identified in the *Arabidopsis thaliana* GenBank database using the tblastn sequence analysis program using default parameters and a P-value cutoff threshold of –4 or –5 or lower, depending on the length of the query sequence. Putative transcription factor sequence hits were then screened to identify those containing particular sequence strings. If the sequence hits contained such sequence strings, the sequences were confirmed as transcription factors.

Alternatively, *Arabidopsis thaliana* cDNA libraries derived from different tissues or treatments, or genomic libraries were screened to identify novel members of a transcription family using a low stringency hybridization approach. Probes were synthesized using gene specific primers in a standard PCR reaction (annealing temperature 60° C) and labeled with ^{32}P dCTP using the High Prime DNA Labeling Kit (Boehringer Mannheim Corp. (now Roche Diagnostics Corp., Indianapolis, IN). Purified radiolabelled probes were added to filters immersed in Church hybridization medium (0.5 M $NaPO_4$ pH 7.0, 7% SDS, 1% w/v bovine serum albumin) and hybridized overnight at 60°C with shaking. Filters were washed two times for 45 to 60 minutes with 1xSCC, 1% SDS at 60° C.

To identify additional sequence 5' or 3' of a partial cDNA sequence in a cDNA library, 5' and 3' rapid amplification of cDNA ends (RACE) was performed using the MARATHON cDNA amplification kit (Clontech, Palo Alto, CA). Generally, the method entailed first isolating poly(A) mRNA, performing first and second strand cDNA synthesis to generate double stranded cDNA, blunting cDNA ends, followed by ligation of the MARATHON Adaptor to the cDNA to form a library of adaptor-ligated ds cDNA.

Gene-specific primers were designed to be used along with adaptor specific primers for both 5' and 3' RACE reactions. Nested primers, rather than single primers, were used to increase PCR specificity. Using 5' and 3' RACE reactions, 5' and 3' RACE fragments were obtained, sequenced and cloned. The process can be repeated until 5' and 3' ends of the full-length gene were identified. Then the full-length cDNA was generated by PCR using primers specific to 5' and 3' ends of the gene by end-to-end PCR.

Example II: Construction of Expression Vectors

For the experiments in Example XIII, two types of constructs were used to modulate the activity of lead transcription factors and test the activity of orthologs and paralogs in transgenic plants. These included direct promoter fusion constructs and two component transformation systems.

For direct promoter fusion, expression of a single full-length wild-type version of a transcription factor polynucleotide sequence was driven by fusing the polynucleotide directly to a promoter. A number of different promoters may be used, such as the native promoter or that gene, or a promoter that drives tissue specific or conditional expression. In the direct promoter fusion assays found in Example XIII, the CaMV 35S constitutive promoter was used. To clone the sequence into the vector, both pMEN20, derived from pMON316 (Sanders et al. (1987) *Nucleic Acids Res.* 15:1543-1558), and the amplified DNA fragment (the sequence was amplified from a genomic or cDNA library using primers specific to sequences upstream and downstream of the coding region) were digested separately with SalI and NotI restriction enzymes at 37° C for 2 hours. The digestion products were subject to electrophoresis in a 0.8% agarose gel and visualized by ethidium bromide staining. The DNA fragments containing the sequence and the linearized plasmid were excised and purified by using a QIAQUICK gel extraction kit (Qiagen, Valencia CA). The fragments of interest were ligated at a ratio of 3:1 (vector to insert). Ligation reactions

using T4 DNA ligase (New England Biolabs, Beverly MA) were carried out at 16° C for 16 hours. The ligated DNAs were transformed into competent cells of the *E. coli* strain DH5alpha by using the heat shock method. The transformations were plated on LB plates containing 50 mg/l kanamycin (Sigma Chemical Co. St. Louis MO). Individual colonies were grown overnight in five milliliters of LB broth containing 50 mg/l kanamycin at 37° C. Plasmid DNA was purified by using Qiaquick Mini Prep kits (Qiagen).

For two component supertransformation (2 comp. supTfn), two separate constructs were used: Promoter::LexA-GAL4TA and opLexA::TF. The first of these (Promoter::LexA-GAL4TA) comprised a desired promoter cloned in front of a LexA DNA binding domain fused to a GAL4 activation domain. The construct vector backbone (pMEN48; P5375) also carried a kanamycin resistance marker, along with an opLexA::GFP reporter. Transgenic lines were obtained containing this first component, and a line was selected that showed reproducible expression of the reporter gene in the desired pattern through a number of generations. A homozygous population was established for that line, and the population was supertransformed with the second construct (opLexA::TF) carrying the transcription factor of interest cloned behind a LexA operator site. This second construct vector backbone (pMEN53; P5381) also contained a sulfonamide resistance marker.

Example III: Transformation of *Agrobacterium* with the Expression Vector

After the plasmid vector containing the gene was constructed, the vector was used to transform *Agrobacterium tumefaciens* cells expressing the gene products. The stock of *Agrobacterium tumefaciens* cells for transformation was made as described by Nagel et al. (1990) *FEMS Microbiol Letts.* 67: 325-328. *Agrobacterium* strain ABI was grown in 250 ml LB medium (Sigma) overnight at 28°C with shaking until an absorbance over 1 cm at 600 nm (A_{600}) of 0.5 – 1.0 was reached. Cells were harvested by centrifugation at 4,000 x g for 15 min at 4°C. Cells were then resuspended in 250 µl chilled buffer (1 mM HEPES, pH adjusted to 7.0 with KOH). Cells were centrifuged again as described above and resuspended in 125 µl chilled buffer. Cells were then centrifuged and resuspended two more times in the same HEPES buffer as described above at a volume of 100 µl and 750 µl, respectively. Resuspended cells were then distributed into 40 µl aliquots, quickly frozen in liquid nitrogen, and stored at -80° C.

Agrobacterium cells were transformed with plasmids prepared as described above following the protocol described by Nagel et al. (*supra*). For each DNA construct to be transformed, 50 – 100 ng DNA (generally resuspended in 10 mM Tris-HCl, 1 mM EDTA, pH 8.0) was mixed with 40 µl of *Agrobacterium* cells. The DNA/cell mixture was then transferred to a chilled cuvette with a 2mm electrode gap and subject to a 2.5 kV charge dissipated at 25 µF and 200 µF using a Gene Pulser II apparatus (Bio-Rad, Hercules, CA). After electroporation, cells were immediately resuspended in 1.0 ml LB and allowed to recover without antibiotic selection for 2 – 4 hours at 28° C in a shaking incubator. After recovery, cells were plated onto selective medium of LB broth containing 100 µg/ml spectinomycin

(Sigma) and incubated for 24-48 hours at 28° C. Single colonies were then picked and inoculated in fresh medium. The presence of the plasmid construct was verified by PCR amplification and sequence analysis.

Example IV: Transformation of *Arabidopsis* Plants with *Agrobacterium tumefaciens* with Expression Vector

After transformation of *Agrobacterium tumefaciens* with plasmid vectors containing the gene, single *Agrobacterium* colonies were identified, propagated, and used to transform *Arabidopsis* plants. Briefly, 500 ml cultures of LB medium containing 50 mg/l kanamycin were inoculated with the colonies and grown at 28° C with shaking for 2 days until an optical absorbance at 600 nm wavelength over 1 cm (A_{600}) of > 2.0 is reached. Cells were then harvested by centrifugation at 4,000 x g for 10 min, and resuspended in infiltration medium (1/2 X Murashige and Skoog salts (Sigma), 1 X Gamborg's B-5 vitamins (Sigma), 5.0% (w/v) sucrose (Sigma), 0.044 μM benzylamino purine (Sigma), 200 μl/l Silwet L-77 (Lehle Seeds) until an A_{600} of 0.8 was reached.

Prior to transformation, *Arabidopsis thaliana* seeds (ecotype Columbia) were sown at a density of ~10 plants per 4" pot onto Pro-Mix BX potting medium (Hummert International) covered with fiberglass mesh (18 mm X 16 mm). Plants were grown under continuous illumination (50-75 $\mu E/m^2/sec$) at 22-23° C with 65-70% relative humidity. After about 4 weeks, primary inflorescence stems (bolts) are cut off to encourage growth of multiple secondary bolts. After flowering of the mature secondary bolts, plants were prepared for transformation by removal of all siliques and opened flowers.

The pots were then immersed upside down in the mixture of *Agrobacterium* infiltration medium as described above for 30 sec, and placed on their sides to allow draining into a 1' x 2' flat surface covered with plastic wrap. After 24 h, the plastic wrap was removed and pots are turned upright. The immersion procedure was repeated one week later, for a total of two immersions per pot. Seeds were then collected from each transformation pot and analyzed following the protocol described below.

Example V: Identification of *Arabidopsis* Primary Transformants

Seeds collected from the transformation pots were sterilized essentially as follows. Seeds were dispersed into in a solution containing 0.1% (v/v) Triton X-100 (Sigma) and sterile water and washed by shaking the suspension for 20 min. The wash solution was then drained and replaced with fresh wash solution to wash the seeds for 20 min with shaking. After removal of the ethanol/detergent solution, a solution containing 0.1% (v/v) Triton X-100 and 30% (v/v) bleach (CLOROX; Clorox Corp. Oakland CA) was added to the seeds, and the suspension was shaken for 10 min. After removal of the bleach/detergent solution, seeds were then washed five times in sterile distilled water. The seeds were stored in the last wash water at 4° C for 2 days in the dark before being plated onto antibiotic selection medium (1 X Murashige and Skoog salts (pH adjusted to 5.7 with 1M KOH), 1 X Gamborg's B-5 vitamins, 0.9% phytagar (Life Technologies), and 50 mg/l kanamycin). Seeds were germinated under

continuous illumination (50-75 $\mu E/m^2/sec$) at 22-23°C. After 7-10 days of growth under these conditions, kanamycin resistant primary transformants (T_1 generation) were visible and obtained. These seedlings were transferred first to fresh selection plates where the seedlings continued to grow for 3-5 more days, and then to soil (Pro-Mix BX potting medium).

Primary transformants were crossed and progeny seeds (T_2) collected; kanamycin resistant seedlings were selected and analyzed. The expression levels of the recombinant polynucleotides in the transformants vary from about a 5% expression level increase to a least a 100% expression level increase. Similar observations are made with respect to polypeptide level expression.

Example VI: Identification of *Arabidopsis* Plants with Transcription Factor Gene Knockouts

The screening of insertion mutagenized *Arabidopsis* collections for null mutants in a known target gene was essentially as described in Krysan et al. (1999) *Plant Cell* 11: 2283-2290. Briefly, gene-specific primers, nested by 5-250 base pairs to each other, were designed from the 5' and 3' regions of a known target gene. Similarly, nested sets of primers were also created specific to each of the T-DNA or transposon ends (the "right" and "left" borders). All possible combinations of gene specific and T-DNA/transposon primers were used to detect by PCR an insertion event within or close to the target gene. The amplified DNA fragments were then sequenced which allows the precise determination of the T-DNA/transposon insertion point relative to the target gene. Insertion events within the coding or intervening sequence of the genes were deconvoluted from a pool comprising a plurality of insertion events to a single unique mutant plant for functional characterization. The method is described in more detail in Yu and Adam, US Application Serial No. 09/177,733 filed October 23, 1998.

Example VII: Identification of Modified Phenotypes in Overexpression or Gene Knockout Plants.

Experiments were performed to identify those transformants or knockouts that exhibited modified biochemical characteristics. Among the biochemicals that were assayed were insoluble sugars, such as arabinose, fucose, galactose, mannose, rhamnose or xylose or the like; prenyl lipids, such as lutein, beta-carotene, xanthophyll-1, xanthophyll-2, chlorophylls A or B, or alpha-, delta- or gamma-tocopherol or the like; fatty acids, such as 16:0 (palmitic acid), 16:1 (palmitoleic acid), 18:0 (stearic acid), 18:1 (oleic acid), 18:2 (linoleic acid), 20:0 , 18:3 (linolenic acid), 20:1 (eicosenoic acid), 20:2, 22:1 (erucic acid) or the like; waxes, such as by altering the levels of C29, C31, or C33 alkanes; sterols, such as brassicasterol, campesterol, stigmasterol, sitosterol or stigmastanol or the like, glucosinolates, protein or oil levels.

Fatty acids were measured using two methods depending on whether the tissue was from leaves or seeds. For leaves, lipids were extracted and esterified with hot methanolic H_2SO_4 and partitioned into hexane from methanolic brine. For seed fatty acids, seeds were pulverized and extracted in methanol:heptane:toluene:2,2-dimethoxypropane:H_2SO_4 (39:34:20:5:2) for 90 minutes at 80°C. After cooling to room temperature the upper phase, containing the seed fatty acid esters, was subjected to GC

analysis. Fatty acid esters from both seed and leaf tissues were analyzed with a SUPELCO SP-2330 column (Supelco, Bellefonte, PA).

Glucosinolates were purified from seeds or leaves by first heating the tissue at 95°C for 10 minutes. Preheated ethanol:water (50:50) is added and after heating at 95°C for a further 10 minutes, the extraction solvent is applied to a DEAE Sephadex column (Pharmacia) which had been previously equilibrated with 0.5 M pyridine acetate. Desulfoglucosinolates were eluted with 300 ul water and analyzed by reverse phase HPLC monitoring at 226 nm.

For wax alkanes, samples were extracted using an identical method as fatty acids and extracts were analyzed on a HP 5890 GC coupled with a 5973 MSD. Samples were chromatographically isolated on a J&W DB35 mass spectrometer (J&W Scientific Agilent Technologies, Folsom, CA).

To measure prenyl lipid levels, seeds or leaves were pulverized with 1 to 2% pyrogallol as an antioxidant. For seeds, extracted samples were filtered and a portion removed for tocopherol and carotenoid/chlorophyll analysis by HPLC. The remaining material was saponified for sterol determination. For leaves, an aliquot was removed and diluted with methanol and chlorophyll A, chlorophyll B, and total carotenoids measured by spectrophotometry by determining optical absorbance at 665.2 nm, 652.5 nm, and 470 nm. An aliquot was removed for tocopherol and carotenoid/chlorophyll composition by HPLC using a Waters µBondapak C18 column (4.6 mm x 150 mm). The remaining methanolic solution was saponified with 10% KOH at 80° C for one hour. The samples were cooled and diluted with a mixture of methanol and water. A solution of 2% methylene chloride in hexane was mixed in and the samples were centrifuged. The aqueous methanol phase was again re-extracted 2% methylene chloride in hexane and, after centrifugation, the two upper phases were combined and evaporated. 2% methylene chloride in hexane was added to the tubes and the samples were then extracted with one ml of water. The upper phase was removed, dried, and resuspended in 400 ul of 2% methylene chloride in hexane and analyzed by gas chromatography using a 50 m DB-5ms (0.25 mm ID, 0.25 um phase, J&W Scientific).

Insoluble sugar levels were measured by the method essentially described by Reiter et al. (1999), *Plant J.* 12: 335-345. This method analyzes the neutral sugar composition of cell wall polymers found in *Arabidopsis* leaves. Soluble sugars were separated from sugar polymers by extracting leaves with hot 70% ethanol. The remaining residue containing the insoluble polysaccharides was then acid hydrolyzed with allose added as an internal standard. Sugar monomers generated by the hydrolysis were then reduced to the corresponding alditols by treatment with $NaBH_4$, then were acetylated to generate the volatile alditol acetates which were then analyzed by GC-FID. Identity of the peaks was determined by comparing the retention times of known sugars converted to the corresponding alditol acetates with the retention times of peaks from wild-type plant extracts. Alditol acetates were analyzed on a Supelco SP-2330 capillary column (30 m x 250 µm x 0.2 µm) using a temperature program beginning at 180° C for 2 minutes followed by an increase to 220° C in 4 minutes. After holding at 220° C for 10 minutes, the oven

temperature is increased to 240° C in 2 minutes and held at this temperature for 10 minutes and brought back to room temperature.

To identify plants with alterations in total seed oil or protein content, 150mg of seeds from T2 progeny plants were subjected to analysis by Near Infrared Reflectance Spectroscopy (NIRS) using a Foss NirSystems Model 6500 with a spinning cup transport system. NIRS is a non-destructive analytical method used to determine seed oil and protein composition. Infrared is the region of the electromagnetic spectrum located after the visible region in the direction of longer wavelengths. 'Near infrared' owns its name for being the infrared region near to the visible region of the electromagnetic spectrum. For practical purposes, near infrared comprises wavelengths between 800 and 2500 nm. NIRS is applied to organic compounds rich in O-H bonds (such as moisture, carbohydrates, and fats), C-H bonds (such as organic compounds and petroleum derivatives), and N-H bonds (such as proteins and amino acids). The NIRS analytical instruments operate by statistically correlating NIRS signals at several wavelengths with the characteristic or property intended to be measured. All biological substances contain thousands of C-H, O-H, and N-H bonds. Therefore, the exposure to near infrared radiation of a biological sample, such as a seed, results in a complex spectrum which contains qualitative and quantitative information about the physical and chemical composition of that sample.

The numerical value of a specific analyte in the sample, such as protein content or oil content, is mediated by a calibration approach known as chemometrics. Chemometrics applies statistical methods such as multiple linear regression (MLR), partial least squares (PLS), and principle component analysis (PCA) to the spectral data and correlates them with a physical property or other factor, that property or factor is directly determined rather than the analyte concentration itself. The method first provides "wet chemistry" data of the samples required to develop the calibration.

Calibration of NIRS response was performed using data obtained by wet chemical analysis of a population of *Arabidopsis* ecotypes that were expected to represent diversity of oil and protein levels.

The exact oil composition of each ecotype used in the calibration experiment was performed using gravimetric analysis of oils extracted from seed samples (0.5 g or 1.0 g) by the accelerated solvent extraction method (ASE; Dionex Corp, Sunnyvale, CA). The extraction method was validated against certified canola samples (Community Bureau of Reference, Belgium). Seed samples from each ecotype (0.5 g or 1g) were subjected to accelerated solvent extraction and the resulting extracted oil weights compared to the weight of oil recovered from canola seed that has been certified for oil content (Community Bureau of Reference). The oil calibration equation was based on 57 samples with a range of oil contents from 27.0 % to 50.8 %. To check the validity of the calibration curve, an additional set of samples was extracted by ASE and predicted using the oil calibration equation. This validation set counted 46 samples, ranging from 27.9 % to 47.5 % oil, and had a predicted standard error of performance of 0.63 %. The wet chemical method for protein was elemental analysis (%N X 6.0) using the average of 3 representative samples of 5 mg each validated against certified ground corn (NIST). The

instrumentation was an Elementar Vario-EL III elemental analyzer operated in CNS operating mode (Elementar Analysensysteme GmbH, Hanau, Germany).

The protein calibration equation was based on a library of 63 samples with a range of protein contents from 17.4 % to 31.2 %. An additional set of samples was analyzed for protein by elemental analysis (n = 57) and scanned by NIRS in order to validate the protein prediction equation. The protein range of the validation set was from 16.8 % to 31.2 % and the standard error of prediction was 0.468 %.

NIRS analysis of *Arabidopsis* seed was carried out on between 40-300 mg experimental sample. The oil and protein contents were predicted using the respective calibration equations.

Data obtained from NIRS analysis was analyzed statistically using a nearest-neighbor (N-N) analysis. The N-N analysis allows removal of within-block spatial variability in a fairly flexible fashion, which does not require prior knowledge of the pattern of variability in the chamber. Ideally, all hybrids are grown under identical experimental conditions within a block (rep). In reality, even in many block designs, significant within-block variability exists. Nearest-neighbor procedures are based on assumption that environmental effect of a plot is closely related to that of its neighbors. Nearest-neighbor methods use information from adjacent plots to adjust for within-block heterogeneity and so provide more precise estimates of treatment means and differences. If there is within-plot heterogeneity on a spatial scale that is larger than a single plot and smaller than the entire block, then yields from adjacent plots will be positively correlated. Information from neighboring plots can be used to reduce or remove the unwanted effect of the spatial heterogeneity, and hence improve the estimate of the treatment effect. Data from neighboring plots can also be used to reduce the influence of competition between adjacent plots. The Papadakis N-N analysis can be used with designs to remove within-block variability that would not be removed with the standard split plot analysis (Papadakis (1973) Inst. d'Amelior. Plantes Thessaloniki (Greece) Bull. *Scientif.* No. 23; Papadakis (1984) *Proc. Acad. Athens* 59: 326-342.

Experiments were performed to identify those transformants or knockouts that exhibited modified sugar-sensing. For such studies, seeds from transformants were germinated on media containing 5% glucose or 9.4% sucrose which normally partially restrict hypocotyl elongation. Plants with altered sugar sensing may have either longer or shorter hypocotyls than normal plants when grown on this media. Additionally, other plant traits may be varied such as root mass.

Experiments may be performed to identify those transformants or knockouts that exhibited an improved pathogen tolerance. For such studies, the transformants are exposed to biotropic fungal pathogens, such as *Erysiphe orontii*, and necrotropic fungal pathogens, such as *Fusarium oxysporum*. *Fusarium oxysporum* isolates cause vascular wilts and damping off of various annual vegetables, perennials and weeds (Mauch-Mani and Slusarenko (1994) *Molec Plant-Microbe Interact.* 7: 378-383). For *Fusarium oxysporum* experiments, plants are grown on Petri dishes and sprayed with a fresh spore suspension of *F. oxysporum*. The spore suspension is prepared as follows: A plug of fungal hyphae from a plate culture is placed on a fresh potato dextrose agar plate and allowed to spread for one week. Five ml sterile water is then added to the plate, swirled, and pipetted into 50 ml Armstrong *Fusarium* medium.

Spores are grown overnight in *Fusarium* medium and then sprayed onto plants using a Preval paint sprayer. Plant tissue is harvested and frozen in liquid nitrogen 48 hours post-infection.

Erysiphe orontii is a causal agent of powdery mildew. For *Erysiphe orontii* experiments, plants are grown approximately 4 weeks in a greenhouse under 12 hour light (20°C, ~30% relative humidity (rh)). Individual leaves are infected with *E. orontii* spores from infected plants using a camel's hair brush, and the plants are transferred to a Percival growth chamber (20°C, 80% rh.). Plant tissue is harvested and frozen in liquid nitrogen 7 days post-infection.

Botrytis cinerea is a necrotrophic pathogen. *Botrytis cinerea* is grown on potato dextrose agar under 12 hour light (20°C, ~30% relative humidity (rh)). A spore culture is made by spreading 10 ml of sterile water on the fungus plate, swirling and transferring spores to 10 ml of sterile water. The spore inoculum (approx. 105 spores/ml) is then used to spray 10 day-old seedlings grown under sterile conditions on MS (minus sucrose) media. Symptoms are evaluated every day up to approximately 1 week.

Sclerotinia sclerotiorum hyphal cultures are grown in potato dextrose broth. One gram of hyphae is ground, filtered, spun down and resuspended in sterile water. A 1:10 dilution is used to spray 10 day-old seedlings grown aseptically under a 12 hour light/dark regime on MS (minus sucrose) media. Symptoms are evaluated every day up to approximately 1 week.

Pseudomonas syringae pv maculicola (Psm) strain 4326 and pv maculicola strain 4326 was inoculated by hand at two doses. Two inoculation doses allows the differentiation between plants with enhanced susceptibility and plants with enhanced resistance to the pathogen. Plants are grown for 3 weeks in the greenhouse, then transferred to the growth chamber for the remainder of their growth. Psm ES4326 may be hand inoculated with 1 ml syringe on 3 fully-expanded leaves per plant (4 1/2 wk old), using at least 9 plants per overexpressing line at two inoculation doses, OD=0.005 and OD=0.0005. Disease scoring is performed at day 3 post-inoculation with pictures of the plants and leaves taken in parallel.

In some instances, expression patterns of the pathogen-induced genes (such as defense genes) may be monitored by microarray expériments. In these experiments, cDNAs are generated by PCR and resuspended at a final concentration of ~ 100 ng/µl in 3X SSC or 150mM Na-phosphate (Eisen and Brown (1999) *Methods Enzymol.* 303: 179-205). The cDNAs are spotted on microscope glass slides coated with polylysine. The prepared cDNAs are aliquoted into 384 well plates and spotted on the slides using, for example, an x-y-z gantry (OmniGrid) which may be purchased from GeneMachines (Menlo Park, CA) outfitted with quill type pins which may be purchased from Telechem International (Sunnyvale, CA). After spotting, the arrays are cured for a minimum of one week at room temperature, rehydrated and blocked following the protocol recommended by Eisen and Brown (1999; supra).

Sample total RNA (10 µg) samples are labeled using fluorescent Cy3 and Cy5 dyes. Labeled samples are resuspended in 4X SSC/0.03% SDS/4 µg salmon sperm DNA/2 µg tRNA/ 50mM Na-pyrophosphate, heated for 95°C for 2.5 minutes, spun down and placed on the array. The array is then

covered with a glass coverslip and placed in a sealed chamber. The chamber is then kept in a water bath at 62°C overnight. The arrays are washed as described in Eisen and Brown (1999, *supra*) and scanned on a General Scanning 3000 laser scanner. The resulting files are subsequently quantified using IMAGENE, software (BioDiscovery, Los Angeles CA).

RT-PCR experiments may be performed to identify those genes induced after exposure to biotropic fungal pathogens, such as *Erysiphe orontii*, necrotropic fungal pathogens, such as *Fusarium oxysporum*, bacteria, viruses and salicylic acid, the latter being involved in a nonspecific resistance response in *Arabidopsis thaliana*. Generally, the gene expression patterns from ground plant leaf tissue is examined.

Reverse transcriptase PCR was conducted using gene specific primers within the coding region for each sequence identified. The primers were designed near the 3' region of each DNA binding sequence initially identified.

Total RNA from these ground leaf tissues was isolated using the CTAB extraction protocol. Once extracted total RNA was normalized in concentration across all the tissue types to ensure that the PCR reaction for each tissue received the same amount of cDNA template using the 28S band as reference. Poly(A+) RNA was purified using a modified protocol from the Qiagen OLIGOTEX purification kit batch protocol. cDNA was synthesized using standard protocols. After the first strand cDNA synthesis, primers for Actin 2 were used to normalize the concentration of cDNA across the tissue types. Actin 2 is found to be constitutively expressed in fairly equal levels across the tissue types being investigated.

For RT PCR, cDNA template was mixed with corresponding primers and Taq DNA polymerase. Each reaction consisted of 0.2 µl cDNA template, 2 µl 10X Tricine buffer, 2 µl 10X Tricine buffer and 16.8 µl water, 0.05 µl Primer 1, 0.05 µl, Primer 2, 0.3 µl Taq DNA polymerase and 8.6 µl water.

The 96 well plate is covered with microfilm and set in the thermocycler to start the reaction cycle. By way of illustration, the reaction cycle may comprise the following steps:

STEP 1: 93° C FOR 3 MIN;

Step 2: 93° C for 30 sec;

Step 3: 65° C for 1 min;

Step 4: 72° C for 2 min;

Steps 2, 3 and 4 are repeated for 28 cycles;

Step 5: 72° C for 5 min; and

Step 6 4° C.

To amplify more products, for example, to identify genes that have very low expression, additional steps may be performed: The following method illustrates a method that may be used in this regard. The PCR plate is placed back in the thermocycler for 8 more cycles of steps 2-4.

Step 2 93° C for 30 sec;

Step 3 65° C for 1 min;

Step 4 72° C for 2 min, repeated for 8 cycles; and

Step 5 4° C.

Eight microliters of PCR product and 1.5 µl of loading dye are loaded on a 1.2% agarose gel for analysis after 28 cycles and 36 cycles. Expression levels of specific transcripts are considered low if they were only detectable after 36 cycles of PCR. Expression levels are considered medium or high depending on the levels of transcript compared with observed transcript levels for an internal control such as actin2. Transcript levels are determined in repeat experiments and compared to transcript levels in control (e.g., non-transformed) plants.

Example VIII: Identification of Homologous Sequences

This example describes identification of genes that are orthologous to *Arabidopsis thaliana* transcription factors from a computer homology search.

Homologous sequences, including those of paralogs and orthologs from *Arabidopsis* and other plant species, were identified using database sequence search tools, such as the Basic Local Alignment Search Tool (BLAST) (Altschul et al. (1990) *J. Mol. Biol.* 215: 403-410; and Altschul et al. (1997) *Nucleic* Acid Res. 25: 3389-3402). The tblastx sequence analysis programs were employed using the BLOSUM-62 scoring matrix (Henikoff and Henikoff (1992) *Proc .Natl. Acad. Sci.* 89: 10915-10919). The entire NCBI GenBank database was filtered for sequences from all plants except *Arabidopsis thaliana* by selecting all entries in the NCBI GenBank database associated with NCBI taxonomic ID 33090 (Viridiplantae; all plants) and excluding entries associated with taxonomic ID 3701 (*Arabidopsis thaliana*).

These sequences are compared to sequences representing genes of SEQ ID NO: 2N - 1, wherein N = 1- 229, using the Washington University TBLASTX algorithm (version 2.0a19MP) at the default settings using gapped alignments with the filter "off". For each gene of SEQ ID NO: 2N - 1, wherein N = 1- 229, individual comparisons were ordered by probability score (P-value), where the score reflects the probability that a particular alignment occurred by chance. For example, a score of 3.6e-40 is 3.6 x 10-40. In addition to P-values, comparisons were also scored by percentage identity. Percentage identity reflects the degree to which two segments of DNA or protein are identical over a particular length. Examples of sequences so identified are presented in Table 10 and Table 12. Paralogous or orthologous sequences were readily identified and available in GenBank by Accession number (Table 10; Test sequence ID). The percent sequence identity among these sequences can be as low as 47%, or even lower sequence identity.

Candidate paralogous sequences were identified among *Arabidopsis* transcription factors through alignment, identity, and phylogenic relationships. A list of paralogs is shown in Table 12. Candidate orthologous sequences were identified from proprietary unigene sets of plant gene sequences in *Zea mays*, *Glycine max* and *Oryza sativa* based on significant homology to *Arabidopsis* transcription factors.

These candidates were reciprocally compared to the set of *Arabidopsis* transcription factors. If the candidate showed maximal similarity in the protein domain to the eliciting transcription factor or to a paralog of the eliciting transcription factor, then it was considered to be an ortholog. Identified non-*Arabidopsis* sequences that were shown in this manner to be orthologous to the *Arabidopsis* sequences are provided in Table 10.

Example IX: Identification of Orthologous and Paralogous Sequences

Orthologs to *Arabidopsis* genes may identified by several methods, including experimental methods such as hybridization and/or amplification. This example describes how one may identify equivalogs to the *Arabidopsis* AP2 family transcription factor CBF1 (polynucleotide SEQ ID NO: 1955, encoded polypeptide SEQ ID NO: 1956), which confers tolerance to abiotic stresses (Thomashow et al. (2002) U.S. Patent No. 6,417,428), and an example to confirm the function of homologous sequences. In this example, orthologs to CBF1 were found in canola (*Brassica napus*) using polymerase chain reaction (PCR).

Degenerate primers were designed for regions of AP2 binding domain and outside of the AP2 (carboxyl terminal domain):

Mol 368 (reverse) 5'- CAY CCN ATH TAY MGN GGN GT -3' (SEQ ID NO: 2205)

Mol 378 (forward) 5'- GGN ARN ARC ATN CCY TCN GCC -3' (SEQ ID NO: 2206)

(Y: C/T, N: A/C/G/T, H: A/C/T, M: A/C, R: A/G)

Primer Mol 368 is in the AP2 binding domain of CBF1 (amino acid sequence: His-Pro-Ile-Tyr-Arg-Gly-Val) while primer Mol 378 is outside the AP2 domain (carboxyl terminal domain) (amino acid sequence: Met-Ala-Glu-Gly-Met-Leu-Leu-Pro).

The genomic DNA isolated from *B. napus* was PCR-amplified by using these primers following these conditions: an initial denaturation step of 2 min at 93° C; 35 cycles of 93° C for 1 min, 55° C for 1 min, and 72° C for 1 min ; and a final incubation of 7 min at 72° C at the end of cycling.

The PCR products were separated by electrophoresis on a 1.2% agarose gel and transferred to nylon membrane and hybridized with the AT CBF1 probe prepared from *Arabidopsis* genomic DNA by PCR amplification. The hybridized products were visualized by colorimetric detection system (Boehringer Mannheim) and the corresponding bands from a similar agarose gel were isolated using the Qiagen Extraction Kit (Qiagen). The DNA fragments were ligated into the TA clone vector from TOPO TA Cloning Kit (Invitrogen) and transformed into *E. coli* strain TOP10 (Invitrogen).

Seven colonies were picked and the inserts were sequenced on an ABI 377 machine from both strands of sense and antisense after plasmid DNA isolation. The DNA sequence was edited by sequencer and aligned with the AtCBF1 by GCG software and NCBI blast searching.

The nucleic acid sequence and amino acid sequence of one canola ortholog found in this manner (bnCBF1; polynucleotide SEQ ID NO: 2203 and polypeptide SEQ ID NO: 2204) identified by this process is shown in the Sequence Listing.

The aligned amino acid sequences show that the bnCBF1 gene has 88% identity with the *Arabidopsis* sequence in the AP2 domain region and 85% identity with the *Arabidopsis* sequence outside the AP2 domain when aligned for two insertion sequences that are outside the AP2 domain.

Similarly, paralogous sequences to *Arabidopsis* genes, such as *CBF1*, may also be identified.

Two paralogs of CBF1 from *Arabidopsis thaliana*: *CBF2* and *CBF3*. *CBF2* and *CBF3* have been cloned and sequenced as described below. The sequences of the DNA SEQ ID NO: 1957 and 1959 and encoded proteins SEQ ID NO: 1958 and 1960 are set forth in the Sequence Listing.

A lambda cDNA library prepared from RNA isolated from *Arabidopsis thaliana* ecotype Columbia (Lin and Thomashow (1992) *Plant Physiol.* 99: 519-525) was screened for recombinant clones that carried inserts related to the *CBF1* gene (Stockinger et al. (1997) *Proc. Natl. Acad. Sci.* 94:1035-1040). CBF1 was ^{32}P-radiolabeled by random priming (Sambrook et al.(1998) *supra*) and used to screen the library by the plaque-lift technique using standard stringent hybridization and wash conditions (Hajela et al. (1990) *Plant Physiol.* 93:1246-1252; Sambrook et al.(1998) *supra*) 6 X SSPE buffer, 60° C for hybridization and 0.1 X SSPE buffer and 60° C for washes). Twelve positively hybridizing clones were obtained and the DNA sequences of the cDNA inserts were determined. The results indicated that the clones fell into three classes. One class carried inserts corresponding to *CBF1*. The two other classes carried sequences corresponding to two different homologs of *CBF1*, designated *CBF2* and *CBF3*. The nucleic acid sequences and predicted protein coding sequences for *Arabidopsis CBF1*, *CBF2* and *CBF3* are listed in the Sequence Listing (SEQ ID NOs: 1955, 1957, 1959 and SEQ ID NOs: 1956, 1958, 1960, respectively). The nucleic acid sequences and predicted protein coding sequence for *Brassica napus* CBF ortholog is listed in the Sequence Listing (SEQ ID NOs: 2203 and 2204, respectively).

A comparison of the nucleic acid sequences of *Arabidopsis CBF1*, *CBF2* and *CBF3* indicate that they are 83 to 85% identical as shown in Table 14.

TABLE 14

	Percent identity[a]	
	DNA[b]	Polypeptide
cbf1/cbf2	85	86
cbf1/cbf3	83	84
cbf2/cbf3	84	85

[a] Percent identity was determined using the *Clustal* algorithm from the Megalign program (DNASTAR, Inc.).

[b] Comparisons of the nucleic acid sequences of the open reading frames are shown.

Similarly, the amino acid sequences of the three CBF polypeptides range from 84 to 86% identity. An alignment of the three amino acidic sequences reveals that most of the differences in amino acid sequence occur in the acidic C-terminal half of the polypeptide. This region of CBF1 serves as an activation domain in both yeast and *Arabidopsis* (not shown).

Residues 47 to 106 of CBF1 correspond to the AP2 domain of the protein, a DNA binding motif that to date, has only been found in plant proteins. A comparison of the AP2 domains of CBF1, CBF2 and CBF3 indicates that there are a few differences in amino acid sequence. These differences in amino acid sequence might have an effect on DNA binding specificity.

Example X: Screen of Plant cDNA library for Sequence Encoding a Transcription Factor DNA Binding Domain That Binds To a Transcription Factor Binding Promoter Element and Demonstration of Protein Transcription Regulation Activity.

The "one-hybrid" strategy (Li and Herskowitz (1993) *Science* 262: 1870-1874) is used to screen for plant cDNA clones encoding a polypeptide comprising a transcription factor DNA binding domain, a conserved domain. In brief, yeast strains are constructed that contain a lacZ reporter gene with either wild-type or mutant transcription factor binding promoter element sequences in place of the normal UAS (upstream activator sequence) of the GALL promoter. Yeast reporter strains are constructed that carry transcription factor binding promoter element sequences as UAS elements are operably linked upstream (5') of a lacZ reporter gene with a minimal GAL1 promoter. The strains are transformed with a plant expression library that contains random cDNA inserts fused to the GAL4 activation domain (GAL4-ACT) and screened for blue colony formation on X-gal-treated filters (X-gal: 5-bromo-4-chloro-3-indolyl-β-D-galactoside; Invitrogen Corporation, Carlsbad CA). Alternatively, the strains are transformed with a cDNA polynucleotide encoding a known transcription factor DNA binding domain polypeptide sequence.

Yeast strains carrying these reporter constructs produce low levels of beta-galactosidase and form white colonies on filters containing X-gal. The reporter strains carrying wild-type transcription factor binding promoter element sequences are transformed with a polynucleotide that encodes a polypeptide comprising a plant transcription factor DNA binding domain operably linked to the acidic activator domain of the yeast GAL4 transcription factor, "GAL4-ACT". The clones that contain a polynucleotide encoding a transcription factor DNA binding domain operably linked to GLA4-ACT can bind upstream of the lacZ reporter genes carrying the wild-type transcription factor binding promoter element sequence, activate transcription of the lacZ gene and result in yeast forming blue colonies on X-gal-treated filters.

Upon screening about 2 x 10^6 yeast transformants, positive cDNA clones are isolated; i.e., clones that cause yeast strains carrying lacZ reporters operably linked to wild-type transcription factor binding promoter elements to form blue colonies on X-gal-treated filters. The cDNA clones do not cause a yeast strain carrying a mutant type transcription factor binding promoter elements fused to LacZ to turn blue.

Thus, a polynucleotide encoding transcription factor DNA binding domain, a conserved domain, is shown to activate transcription of a gene.

Example XI: Gel Shift Assays.

The presence of a transcription factor comprising a DNA binding domain which binds to a DNA transcription factor binding element is evaluated using the following gel shift assay. The transcription factor is recombinantly expressed and isolated from *E. coli* or isolated from plant material. Total soluble protein, including transcription factor, (40 ng) is incubated at room temperature in 10 µl of 1 x binding buffer (15 mM HEPES (pH 7.9), 1 mM EDTA, 30 mM KCl, 5% glycerol, 5% bovine serum albumin, 1 mM DTT) plus 50 ng poly(dl-dC):poly(dl-dC) (Pharmacia, Piscataway NJ) with or without 100 ng competitor DNA. After 10 minutes incubation, probe DNA comprising a DNA transcription factor binding element (1 ng) that has been ^{32}P-labeled by end-filling (Sambrook et al. (1989) *supra*) is added and the mixture incubated for an additional 10 minutes. Samples are loaded onto polyacrylamide gels (4% w/v) and fractionated by electrophoresis at 150V for 2h (Sambrook et al. (1998) supra). The degree of transcription factor-probe DNA binding is visualized using autoradiography. Probes and competitor DNAs are prepared from oligonucleotide inserts ligated into the BamHI site of pUC118 (Vieira et al. (1987) *Methods Enzymol.* 153: 3-11). Orientation and concatenation number of the inserts are determined by dideoxy DNA sequence analysis (Sambrook et al. (1998) supra). Inserts are recovered after restriction digestion with EcoRI and HindIII and fractionation on polyacrylamide gels (12% w/v) (Sambrook et al. (1998) *supra*).

Example XII. Introduction of Polynucleotides into Dicotyledonous Plants

Transcription factor sequences listed in the Sequence Listing recombined into pMEN20 or pMEN65 expression vectors are transformed into a plant for the purpose of modifying plant traits. The cloning vector may be introduced into a variety of cereal plants by means well known in the art such as, for example, direct DNA transfer or *Agrobacterium tumefaciens*-mediated transformation. It is now routine to produce transgenic plants using most dicot plants (see Weissbach and Weissbach, (1989) *supra;* Gelvin et al. (1990) *supra*; Herrera-Estrella et al. (1983) *supra*; Bevan (1984) *supra*; and Klee (1985) *supra*). Methods for analysis of traits are routine in the art and examples are disclosed above.

Example XIII: Examples of Genes that Confer Significant Improvements to Plants

Experiments were performed to identify those transformants or knockouts that exhibited a morphological difference relative to wild-type control plants, i.e., a modified structure and/or development characteristics. For such studies, the transformants were observed by eye to identify novel structural or developmental characteristics associated with the ectopic expression of the polynucleotides or polypeptides of the invention. Examples of genes and equivalogs that confer significant improvements to overexpressing plants are noted below. Experimental observations made with regard to specific genes

whose expression has been modified in overexpressing plants, and potential applications based on these observations, are also presented.

Modified phenotypes observed for particular overexpressor or knockout plants are provided in Table 7. For a particular overexpressor that shows a less beneficial characteristic, it may be more useful to select a plant with a decreased expression of the particular transcription factor. For a particular knockout that shows a less beneficial characteristic, it may be more useful to select a plant with an increased expression of the particular transcription factor.

Table 7 provides exemplary polynucleotide and polypeptide sequences of the invention. The sequences of the Sequence Listing or those in Tables 7 - 11, or those disclosed here, can be used to prepare transgenic plants and plants with altered traits. The specific transgenic plants listed below are produced from the sequences of the Sequence Listing as noted. A number of genes and homologs that confer significant improvements to knockout or overexpressing plants were noted below. Experimental observations made with regard to specific genes whose expression was modified in overexpressing or knockout plants, and potential applications based on these observations, were also presented. As noted in the results of the plate-based physiology assays presented in the tables of this Example, a representative number of sequences from diverse plant species conferred various levels of increased stress tolerance in a range of abiotic stress assays, as noted below. Observed effects of overexpression on flowering time are also noted in the text below. These comparable effects indicate that sequences found within specific clades or subclades are functionally related and can be used to confer various types of abiotic stress tolerance in plants. A number of these genes concurrently confer tolerance to multiple abiotic stresses.

Salt stress assays are intended to find genes that confer better germination, seedling vigor or growth in high salt. Evaporation from the soil surface causes upward water movement and salt accumulation in the upper soil layer where the seeds are placed. Thus, germination normally takes place at a salt concentration much higher than the mean salt concentration of in the whole soil profile. Plants differ in their tolerance to NaCl depending on their stage of development, therefore seed germination, seedling vigor, and plant growth responses are evaluated.

Osmotic stress assays (including NaCl and mannitol assays) are intended to determine if an osmotic stress phenotype is NaCl-specific or if it is a general osmotic stress related phenotype. Plants tolerant to osmotic stress could also have more tolerance to drought and/or freezing.

Drought assays are intended to find genes that mediate better plant survival after short-term, severe water deprivation. Ion leakage will be measured if needed. Osmotic stress tolerance would also support a drought tolerant phenotype.

Temperature stress assays are intended to find genes that confer better germination, seedling vigor or plant growth under temperature stress (cold, freezing and heat).

Sugar sensing assays are intended to find genes involved in sugar sensing by germinating seeds on high concentrations of sucrose and glucose and looking for degrees of hypocotyl elongation. The germination assay on mannitol controls for responses related to osmotic stress. Sugars are key regulatory

molecules that affect diverse processes in higher plants including germination, growth, flowering, senescence, sugar metabolism and photosynthesis. Sucrose is the major transport form of photosynthate and its flux through cells has been shown to affect gene expression and alter storage compound accumulation in seeds (source-sink relationships). Glucose-specific hexose-sensing has also been described in plants and is implicated in cell division and repression of "famine" genes (photosynthetic or glyoxylate cycles).

Germination assays followed modifications of the same basic protocol. Sterile seeds were sown on the conditional media listed below. Plates were incubated at 22° C under 24-hour light (120-130 $\mu Ein/m^2/s$) in a growth chamber. Evaluation of germination and seedling vigor was conducted 3 to 15 days after planting. The basal media was 80% Murashige-Skoog medium (MS) + vitamins.

For salt and osmotic stress germination experiments, the medium was supplemented with 150 mM NaCl or 300 mM mannitol. Growth regulator sensitivity assays were performed in MS media, vitamins, and either 0.3 μM ABA, 9.4% sucrose, or 5% glucose.

Temperature stress cold germination experiments were carried out at 8 °C. Heat stress germination experiments were conducted at 32 °C to 37° C for 6 hours of exposure.

For stress experiments conducted with more mature plants, seeds were germinated and grown for seven days on MS + vitamins + 1% sucrose at 22 °C and then transferred to chilling and heat stress conditions. The plants were either exposed to chilling stress (6 hour exposure to 4-8° C), or heat stress (32° C was applied for five days, after which the plants were transferred back 22 °C for recovery and evaluated after 5 days relative to controls not exposed to the depressed or elevated temperature).

Soil-based drought screens were performed with *Arabidopsis* plants overexpressing the transcription factors listed in the Sequence Listing, where noted below. Seeds from wild-type *Arabidopsis* plants, or plants overexpressing a polypeptide of the invention, were stratified for three days at 4° C in 0.1% agarose. Fourteen seeds of each overexpressor or wild-type were then sown in three inch clay pots containing a 50:50 mix of vermiculite:perlite topped with a small layer of MetroMix 200 and grown for fifteen days under 24 hr light. Pots containing wild-type and overexpressing seedlings were placed in flats in random order. Drought stress was initiated by placing pots on absorbent paper for seven to eight days. The seedlings were considered to be sufficiently stressed when the majority of the pots containing wild-type seedlings within a flat had become severely wilted. Pots were then re-watered and survival was scored four to seven days later. Plants were ranked against wild-type controls for each of two criteria: tolerance to the drought conditions and recovery (survival) following re-watering

At the end of the initial drought period, each pot was assigned a numeric value score depending on the above criteria. A low value was assigned to plants with an extremely poor appearance (i.e., the plants were uniformly brown) and a high value given to plants that were rated very healthy in appearance (i.e., the plants were all green). After the plants were rewatered and incubated an additional four to seven days, the plants were reevaluated to indicate the degree of recovery from the water deprivation treatment.

An analysis was then conducted to determine which plants best survived water deprivation, identifying the transgenes that consistently conferred drought-tolerant phenotypes and their ability to recover from this treatment. The analysis was performed by comparing overall and within-flat tabulations with a set of statistical models to account for variations between batches. Several measures of survival were tabulated, including: (a) the average proportion of plants surviving relative to wild-type survival within the same flat; (b) the median proportion surviving relative to wild-type survival within the same flat; (c) the overall average survival (taken over all batches, flats, and pots); (d) the overall average survival relative to the overall wild-type survival; and (e) the average visual score of plant health before rewatering.

Flowering time was measured by the number of rosette leaves present when a visible inflorescence of approximately 3 cm is apparent. Rosette and total leaf number on the progeny stem are tightly correlated with the timing of flowering (Koornneef et al. (1991) *Mol. Gen. Genet.* 229: 57-66). The vernalization response was also measured. For vernalization treatments, seeds were sown to MS agar plates, sealed with micropore tape, and placed in a 4°C cold room with low light levels for 6-8 weeks. The plates were then transferred to the growth rooms alongside plates containing freshly sown non-vernalized controls. Rosette leaves were counted when a visible inflorescence of approximately 3 cm was apparent.

Results:

Empty cells found in the tables in this Example indicate that results have not yet been obtained for that particular stress experiment. Abbreviations used in the tables in this example include:

(++) Substantially enhanced performance compared to controls. The phenotype was very consistent and growth was much greater than the normal levels of variability observed for that assay.

(+) Enhanced performance compared to controls. The response was consistently above the normal levels of variability observed for that assay.

(wt) response similar to wild-type; and

(germ.) germination.

G9 (SEQ ID NO: 1949 and 1950)

G9 is a putative paralog of G867, and has been referenced in the public literature as *RAP2.8* and *RAV2* (Okamuro et al. (1997) *Proc. Natl. Acad. Sci. USA* 94: 7076-7081; Kagaya et al. (1999) *Nucleic Acids Res.* 27: 470-478). No genetic analysis of the locus has yet been published.

Experimental Observations

We have previously observed that G9 overexpression enhanced root growth. The aim of the present study was to re-assess 35S::G9 lines and determine whether overexpression of the gene could confer enhanced stress tolerance in a comparable manner to G867. We also sought to test whether use of

a two-component overexpression system would produce any strengthening of the phenotype relative to the use of a 35S direct promoter-fusion.

New overexpression lines have been obtained using both a direct promoter fusion construct and a two component expression system. Lines generated by either of these methods exhibited similar phenotypes and displayed a number of morphological effects that had not been observed during our earlier genomics screens. These included a reduction in overall size, alterations in leaf orientation (which potentially indicated a disruption in circadian control), slow growth, and floral abnormalities relative to controls.

We have tested the 35S::G9 two-component and direct-fusion lines under a variety of plate based treatments. All of the direct fusion lines and most of the two-component lines out-performed controls in a germination assay on sodium chloride plates. In addition, many of these lines also showed positive phenotypes when germinated on sucrose, and ABA, as well as in a growth assay under cold conditions.

It should be emphasized that we have obtained comparable developmental effects as well as a strong enhancement of drought related stress tolerance in 35S::G867 lines and in overexpression lines for the other two putative paralogs, G1930 and G993. The almost identical phenotypic effects observed for the four genes strongly suggest that they are functionally equivalent.

Table 15. G9 35S, Direct promoter-fusion and 2-components-supTfn

Line	Transfor-mation	Germ. in high NaCl	Germ. in high mannitol	Sucrose	ABA	Germ in heat	Germ in cold	Growth in heat	Drought	Chilling
302	Direct-fusion	+	wt	+	wt	wt	wt	wt	wt	+
304	Direct-fusion	+	wt	wt	wt	wt	wt	wt	wt	wt
305	Direct-fusion	++	wt	+	wt	wt	wt	wt	wt	+
306	Direct-fusion	+	wt	wt	wt	wt	wt	wt	wt	wt
307	Direct-fusion	+	wt	wt	wt	wt	wt	wt	wt	wt
310	Direct-fusion	+	wt	+	++	wt	wt	wt	wt	+
311	Direct-fusion	++	wt	+	wt	wt	wt	wt	wt	+
312	Direct-fusion	++	wt	wt	+	wt	wt	wt	wt	+
313	Direct-fusion	+	wt	+	+	wt	wt	wt	wt	+
318	Direct-fusion	++	wt	+	wt	wt	wt	wt	wt	wt
483	2-comp	wt	wt	+	+	wt				
485	2-comp	+	wt	wt	wt	wt				
486	2-comp	wt	wt	wt	wt	wt				
488	2-comp	+	wt	+	++	wt				
489	2-comp	+	wt	wt	++	wt				

490	2-comp	wt	wt	+	++	wt				
491	2-comp	wt	wt	+	++	wt				
493	2-comp	wt	wt	+	++	wt				
494	2-comp	+	wt	wt	++	wt				
498	2-comp	+	wt	+	++	wt				

Potential applications

Based on the results of these abiotic stress assays, G867 and related genes such as G9 are excellent candidate genes for improvement of drought and cold-related stress tolerance in commercial species. The morphological effects associated with their overexpression suggests that tissue-specific or conditional promoters might be used to optimize the utility of these genes.

G19 (SEQ ID NO: 3 and 5)

Published information

G19 belongs to the EREBP subfamily of transcription factors and contains only one AP2 domain. G19 corresponds to the previously described gene RAP2.3 (Okamuro et al. (1997) *Proc. Natl. Acad. Sci.* 94:7076-7081). Close inspection of the *Arabidopsis* cDNA sequences of RAP2.3 (AF003096; Okamuro et al. (1997) *supra*), AtEBP (Y09942; Buttner et al. (1997) *Proc. Natl. Acad. Sci.* 94:5961-5966), and ATCADINP (Z37504) suggests that they may correspond to the same gene (Riechmann et al. (1998) *Biol. Chem.* 379:633-646). G19/RAP2.3 is ubiquitously expressed (Okamuro et al. (1997) *supra*). AtEBP was isolated by virtue of the protein-protein interaction between AtEBP and OBF4, a basic-region leucine zipper transcription factor (Buttner et al. (1997) *supra*). AtEBP expression levels in seedlings were increased after treatment with ethylene (ethephon) (Buttner et al. (1997) *supra*). AtEBP was found to bind to GCC-box containing sequences, like that of the PRB-1b promoter (Buttner et al. (1997) *supra*). It has been suggested that the interaction between AtEBP and OBF4 reflects cross-coupling between EREBP and bZIP transcription factors that might be important in regulating gene expression during the plant defense response (Buttner et al. (1997) *supra*).

Experimental observations

Transgenic plants in which G19 is expressed under the control of the 35S promoter were morphologically similar to control plants. G19 is constitutively expressed in the different tissues examined; however G19 expression was significantly repressed by methyl jasmonate (MeJ) and induced by ACC (this latter result correlates with the previously described increase in G19 expression levels in seedlings after treatment with ethylene (ethephon); Buttner et al. (1997) *supra*). G19 was significantly induced upon infection by the fungal pathogen *Erysiphe orontii*. In addition, G19 overexpressing plants were more tolerant to infection with a moderate dose of *Erysiphe orontii*. G19 overexpressing plants were also tested for their tolerance to two other pathogens, the necrotrophic fungal pathogen *Fusarium*

oxysporum and the bacterial pathogen *Pseudomonas syringae*; the transgenic plants were not found to have altered susceptibility to the pathogens.

Both the jasmonic acid and the ethylene signal transduction pathways were involved in the regulation of the defense response and the wound response, and the two pathways have been found to interact synergistically. The regulation of G19 expression by both hormones, its induction upon *Erysiphe orontii* infection, as well as the preliminary data indicating that increased tolerance to that pathogen was conferred by G19 overexpression, suggest that G19 plays a role in the control of the defense and/or wound response. It would be of interest to test G19 overexpressing plants in insect-plant interaction experiments. The increase in tolerance to *Erysiphe orontii* that is conferred by G19 overexpression can be tested using other races of the pathogen. It would also be of interest to test other pathogens in addition to *Erysiphe orontii*, *Fusarium oxysporum*, and *Pseudomonas syringae*.

Since G19 was expressed at significant levels in a constitutive fashion, similar experiments to those described here can be performed with G19 knockout mutant plants to further elucidate the function of this gene.

Potential applications

G19 or its equivalogs can be used to manipulate the plant defense- wound- or insect-response, as well as the jasmonic acid and ethylene signal transduction pathways themselves.

G22 (SEQ ID NO: 5 and 6)

Published information

G22 was identified in the sequence of BAC T13E15 (gene T13E15.5) by The Institute of Genomic Research (TIGR) as a "TINY transcription factor isolog". G22 belongs to the EREBP subfamily and contains only one AP2 domain, and phylogenetic analyses place G22 relatively close to other EREBP subfamily genes, such as, TINY and ATDL4400C (Riechmann et al. (1998) *Biol. Chem.* 379:633-646).

Experimental observations

G22 was constitutively expressed at medium levels. There appeared to be no phenotypic alteration on plant morphology upon G22 overexpression. Plants ectopically overexpressing G22 were more tolerant to high NaCl containing media in a root growth assay compared with wild-type controls.

Potential applications

G22 or its equivalogs can be used to increase plant tolerance to soil salinity during germination, at the seedling stage, or throughout the plant life cycle. Salt tolerance is a particularly desirable phenotype during the germination stage of a crop plant, which would impact survivability and yield.

G28 (SEQ ID NO: 9 and 10)

Published Information

G28 corresponds to AtERF1 (GenBank accession number AB008103) (Fujimoto et al. (2000) *Plant Cell* 12:393-404). G28 appears as gene AT4g17500 in the annotated sequence of *Arabidopsis* chromosome 4 (AL161546.2).

AtERF1 has been shown to have GCC-box binding activity [some defense-related genes that were induced by ethylene were found to contain a short cis-acting element known as the GCC-box: AGCCGCC (Ohme et al. (1990) *Plant Mol. Biol.* 15:941-946)]. Using transient assays in *Arabidopsis* leaves, AtERF1 was found to be able to act as a GCC-box sequence specific transactivator (Fujimoto et al. (2000) *supra*).

AtERF1 expression has been described to be induced by ethylene (two- to three-fold increase in AtERF1 transcript levels 12 h after ethylene treatment) (Fujimoto et al. (2000) *supra*). In the ein2 mutant, the expression of AtERF1 was not induced by ethylene, suggesting that the ethylene induction of AtERF1 is regulated under the ethylene signaling pathway (Fujimoto et al. (2000) *supra*). AtERF1 expression was also induced by wounding, but not by other abiotic stresses (such as cold, salinity, or drought) (Fujimoto et al. (2000) *supra*).

It has been suggested that AtERFs, in general, may act as transcription factors for stress-responsive genes, and that the GCC-box may act as a cis-regulatory element for biotic and abiotic stress signal transduction in addition to its role as an ethylene responsive element (ERE) (Fujimoto et al. (2000) *supra*), but there is no data available on the physiological functions of AtERF1.

Experimental Observations

The function of G28 was analyzed using transgenic plants in which this gene was expressed under the control of the 35S promoter. G28 overexpressing lines were more tolerant to infection with a moderate dose of the fungal pathogen *Erysiphe orontii*. G28 overexpression did not seem to have detrimental effects on plant growth or vigor, since plants from most of the lines were morphologically wild-type. In addition, no difference was detected between those lines and the corresponding wild-type controls in all the biochemical assays that were performed.

G28 was ubiquitously expressed.

G28 overexpressing lines were also more tolerant to *Sclerotinia sclerotiorum* and *Botrytis cinerea*. In a repeat experiment using individual lines, all three lines analyzed showed tolerance to S. sclerotiorum, and two of the three lines tested were more tolerant to *B cinerea*.

Potential Applications

G28 transgenic plants had an altered response to fungal pathogens, in that those plants were more tolerant to the pathogens. Therefore, G28 or its equivalogs can be used to manipulate the defense response in order to generate pathogen-resistant plants.

G47 (SEQ ID NO: 11 and 12)

Published Information

G47 corresponds to gene T22J18.2 (AAC25505).

Experimental Observations

G47 expression levels can be altered by environmental conditions, in particular reduced by salt and osmotic stresses.

The function of G47 was studied using transgenic plants in which the gene was expressed under the control of the 35S promoter. 35S::G47 plants showed enhanced tolerance to osmotic stress. In a root growth assay on PEG containing media, G47 overexpressing transgenic seedlings were larger and had more root growth compared with the wild-type controls. G47 overexpressors also were significantly more drought tolerant than wild-type control plants in a soil-based assay.

Overexpression of G47 also produced a substantial delay in flowering time and caused a marked change in shoot architecture. 35S::G47 transformants were small at early stages and switched to flowering more than a week later than wild-type controls (continuous light conditions). The inflorescences from these plants appeared thick and fleshy, had reduced apical dominance, and exhibited reduced internode elongation leading to a short compact stature. The branching pattern of the stems also appeared abnormal, with the primary shoot becoming "kinked" at each coflorescence node. Additionally, the plants showed reduced fertility and formed rather small siliques that were borne on short pedicels and held vertically, close against the stem.

Additional alterations were detected in the inflorescence stems of 35S::G47 plants. Stem sections from T2-21 and T2-24 plants were of wider diameter, and had large irregular vascular bundles containing a much greater number of xylem vessels than wild type. Furthermore, some of the xylem vessels within the bundles appeared narrow and were possibly more lignified than were those of controls.

G47 was expressed at higher levels in rosette leaves, and transcripts were detected in other tissues (flower, embryo, silique, and germinating seedling), but not in roots.

Potential Applications

G47 or its equivalogs can be used to manipulate flowering time, to modify plant architecture and stem structure (including development of vascular tissues and lignin content) and to improve plant performance under osmotic stress and drought conditions.

Transcription factor equivalogs that modulate lignin content can be valuable. This modulation can allow the quality of wood used for furniture or construction to be improved. Lignin is energy rich; increasing lignin composition is valuable in raising the energy content of wood used for fuel. Conversely, the pulp and paper industries seek wood with a reduced lignin content. Currently, lignin must be removed in a costly process that involves the use of many polluting chemicals. Consequently,

lignin is a serious barrier to efficient pulp and paper production. In addition to forest biotechnology applications, changing lignin content can increase the palatability of various fruits and vegetables.

G226 (SEQ ID NO: 37 and 38)

Published Information

G226 was identified from the *Arabidopsis* BAC sequence, AC002338, based on its sequence similarity within the conserved domain to other Myb family members in *Arabidopsis*. To date, there is no published information regarding the function of this gene.

Experimental Observations

The function of G226 was analyzed through its ectopic overexpression in plants. G226 overexpressors were more tolerant to low nitrogen and high salt stress. They showed more root growth and possibly more root hairs under conditions of nitrogen limitation compared with wild-type controls. Many plants were glabrous and lacked anthocyanin production when under stress such as growth conditions of low nitrogen and high salt. Several G226 overexpressors were glabrous and produce less anthocyanin under stress; these effects might be due to binding site competition with other Myb family transcription factors involved in these functions and not directly related to the primary function of this gene.

Results from the biochemical analysis of G226 overexpressors suggested that one line had higher amounts of seed protein, which could have been a result of increased nitrogen uptake by these plants.

A microarray experiment was done on a separate G226 overexpressing line. The G226 sequence itself was overexpressed 16-fold above wild type, however, very few changes in other gene expression were observed in this line. On the array, a chlorate/nitrate transporter DNA sequence was induced 2.7-fold over wild type, which could explain the low nitrogen tolerant phenotype of the plants and the increased amounts of seed protein in one of the lines. The same DNA sequence was present several times on the array and in all cases the DNA sequence showed induction, adding more validity to the data. Five other genes/DNA sequences induced but had unknown function. A methyltransferase, a pollen-specific protein, and a zinc binding peroxisomal membrane protein encoding sequences were also induced, however their role in regard to the phenotype of the plants is not known.

Table 16. G226 35S, 2-components supTfn

Line	Germ. in high NaCl	Germ. in high mannitol	Sucrose	ABA	Germ in heat	Germ in cold	Growth in heat	Drought	Chilling	G682-like root morph.
308	wt	wt	++	+	wt	wt	wt	wt	wt	+
309	wt	wt	wt	+	wt	+	wt	wt	+	+
313	wt	wt	++	++	wt	+	wt	wt	+	+
316	wt	wt	++	++	wt	wt	wt	wt	wt	+

318	wt	wt	wt	++	wt	wt	wt	wt	wt	+
381	wt	wt	+	wt	wt	wt	wt	wt	wt	+
383	wt	wt	+	+	wt	wt	wt	wt	wt	+
583	wt	wt	wt	+	wt	wt		wt	wt	+
585	wt	wt	wt	+	wt	wt	wt	wt	wt	+

Potential Applications

The results of these abiotic stress tolerance assays indicate that G682 and related genes such as G226 may be used to improve drought and cold-related stress tolerance in plants.

The utilities of a gene or its equivalogs conferring tolerance to conditions of low nitrogen include: (1) Cost savings to the farmer by reducing the amounts of fertilizer needed; (2) Environmental benefits of reduced fertilizer runoff; (3) Improved yield and stress tolerance. In addition, G226 can be used to increase seed protein amounts and/or composition, which may impact yield as well as the nutritional value and production of various food products.

G682 and related genes such as G226 can be used to alter trichome number and distribution in plants. Trichome glands on the surface of many higher plants produce and secrete exudates, which give protection from the elements and pests such as insects, microbes and herbivores. These exudates may physically immobilize insects and spores, may be insecticidal or antimicrobial or they may allergens or irritants to protect against herbivores. It has also been suggested that trichomes may decrease transpiration by decreasing leaf surface airflow, and by exuding chemicals that protect the leaf from the sun.

The reduction in size that was apparent in these lines suggests that the utility of G226 might be optimized by use of different promoters or protein modifications.

G353 (SEQ ID NO: 59 and 60)

Published Information

G353 was identified in the sequence of P1 clone MMN10, GenBank accession number AB0154751, released by the *Arabidopsis* Genome Initiative. G353 corresponds to RHL41 (Kazuoka et al. (2000) *Plant J.* 24:191-203) and Zat12 (Meissner et al. (1997) *Plant Mol. Biol.* 33:615-624). Transgenic *Arabidopsis* plants over-expressing the RHL41 gene showed an increased tolerance to high-intensity light, and also morphological changes of thicker and dark green leaves. The palisade parenchyma was highly developed in the leaves of the transgenic plants. Anthocyanin content, as well as the chlorophyll content, also increased. Antisense transgenic plants exhibited decreased tolerance to high irradiation. RHL41 protein may play a key role in the acclimatization response to changes in light intensity.

Experimental Observations

G353 was uniformly expressed in all tissues and under all conditions tested in RT-PCR experiments. The highest level of expression was observed in rosette leaves, embryos, and siliques.

The function of this gene was analyzed using transgenic plants in which G353 was expressed under the control of the 35S promoter. Overexpression of G353 in resulted in enhanced tolerance to osmotic stress and drought tolerance in a soil-based assay.

Overexpression also affected flower morphology to a significant degree. 35S::G353 plants had a reduction in flower pedicel length, and downward pointing siliques. This phenotype was very similar to that described for the brevipedicellus (bp) mutant (Koornneef et al. (1983) *J. Hered.* 74:265-272) and in overexpression of a related gene, G354. Other morphological changes in shoots were also observed in 35S::G353 plants. Leaves had short petioles, were rather flat, rounded, and sometimes showed changes in coloration. These effects were observed in varying degrees in the majority of transformants. Severely affected plants were tiny, had contorted leaves, poor fertility, and produced few seeds. Overexpression of G353 in *Arabidopsis* resulted in an increase in seed glucosinolate M39494 in two T2 lines.

Potential Applications

G353 or its equivalogs can be used to alter inflorescence structure, which may have value in production of novel ornamental plants.

G353 or its equivalogs can be used to alter a plant's response to water deficit conditions and, therefore, be used to engineer plants with enhanced tolerance to drought, salt stress, and freezing.

Increases or decreases in specific glucosinolates or total glucosinolate content may be desirable depending upon the particular application. For example: (1) Glucosinolates are undesirable components of the oilseeds used in animal feed, since they produce toxic effects. Low-glucosinolate varieties of canola have been developed to combat this problem; (2) Some glucosinolates have anti-cancer activity; thus, increasing the levels or composition of these compounds might be of interest from a nutraceutical standpoint; (3) Glucosinolates form part of a plants natural defense against insects; modification of glucosinolate composition or quantity could therefore afford increased protection from predators; furthermore, in edible crops, tissue specific promoters might be used to ensure that these compounds accumulate specifically in tissues, such as the epidermis, which are not taken for consumption.

G354 (SEQ ID NO: 61 and 62)

Published Information

G354 was identified in the sequence of BAC clone F12M12, GenBank accession number AL355775, released by the *Arabidopsis* Genome Initiative. G354 corresponds to ZAT7 (Meissner et al. *Plant Mol. Biol.* 33:615-624).

Experimental Observations

Greatest levels of expression of G354 were observed in rosette leaves, embryos, and siliques. Some expression of G354 was also observed in flowers.

The function of this gene was analyzed using transgenic plants in which G353 was overexpressed under the control of the 35S promoter. 35S::G354 plants had a reduction in flower pedicel length, and downward pointing siliques. This phenotype was very similar to that described for the brevipedicellus (bp) mutant (Koornneef et al. (1983) *J. Hered.* 74:265-272) and in overexpression of a related gene, G353. Other morphological changes in shoots were also observed in 35S::G354 plants. Many 35S::G354 seedlings had abnormal cotyledons, elongated, thickened hypocotyls, and short roots. The majority of T1 plants had a very extreme phenotype, were tiny, and arrested development without forming inflorescences. T1 plants showing more moderate effects had poor seed yield.

Overexpression of G354 in *Arabidopsis* resulted in seedlings with an altered response to light. In darkness, G354 seedlings failed to etiolate. The phenotype was most severe in seedlings from one line where overexpression of the transgene resulted in reduced open and greenish cotyledons.

Potential Applications

G354 or its equivalogs can be used to alter inflorescence structure, which may have value in production of novel ornamental plants.

G354 modifies the light response and thus G354 or its equivalogs may be useful for modifying plant growth or development, for example, photomorphogenesis in poor light, or accelerating flowering time in response to various light intensities, quality or duration to which a non-transformed plant would not similarly respond. Elimination of shading responses may lead to increased planting densities with subsequent yield enhancement.

G481 (SEQ ID NO: 87 and 88)

Published Information

G481 is a member of the HAP3 subgroup of the CCAAT-box binding transcription factor family, and is equivalent to AtHAP3a which was identified by Edwards et al. ((1998) *Plant Physiol.* 117:1015-1022) as an EST with extensive sequence homology to the yeast HAP3. Northern blot data from five different tissue samples indicated that G481 was primarily expressed in flower and/or silique, and root tissue.

Experimental Observations

We have now generated additional sets of 35S lines for G481 using the two component system. Two batches of 2-component 35S lines were obtained (lines 301-320 and 741-751). The majority of plants from these T1 sets displayed no consistent difference in morphology to wild-type controls. However, a number of individuals were observed to be late flowering (#305, 310, 315, 744, 748) under the conditions of continuous light.

To confirm the above phenotype, a selection of T2 lines were examined under 24-hour light conditions; late flowering was also observed in that generation in a significant number of these plants, as detailed below:

T2-303: 6/6 plants were slightly late flowering (approximately 2-3 days after controls).

T2-307: 4/6 plants were slightly late flowering (approximately 2-3 days after controls), 2/6 appeared wild type.

T2-309: 4/6 plants were slightly late flowering (approximately 2-3 days after controls), 2/6 appeared wild type.

T2-310: 6/6 plants were moderately late flowering (1-2 weeks after controls).

T2-312: 6/6 plants were moderately late flowering (approximately 1 week after controls).

T2-741: 6/6 plants appeared wild-type.

T2-742: 6/6 plants appeared wild-type.

T2-744: 6/6 plants appeared wild-type.

In addition to analyzing these two component lines, we also re-examined some of the 35S::G481 lines that we had generated during our genomics screens. 35S::G481 line 3 was back-crossed to wild-type and F1 plants were examined. 18/18 of these F1 plants showed a moderate delay in the onset of flowering by about 1-2 weeks.

Of the ten two-component lines submitted for physiological assays, the following showed a segregation on selection plates in the T2 generation that was compatible with the transgene being present at a single locus: 304, 309, 312, 317, 741, 748. Lines 305, 315, 318, 742 showed segregations that were compatible with insertions at multiple loci.

In plate-based physiology assays, tolerance was seen to drought related stress: four of ten two-component lines tested were less sensitive in the ABA germination assay, and two of these lines also displayed cold tolerance in a germination assay.

G481 overexpressing plants were found to be more tolerant to drought in a soil-based assay.

Table 17. G481 35S, 2-components-supertransformation (supTfn)

Line	Germ. in high NaCl	Germ. in high mannitol	Sucrose	ABA	Germ. in heat	Germ. in cold	Growth in heat	Drought	Chilling
304	wt	wt	wt	+	wt	wt	wt	wt	wt
305	wt	wt	wt	wt	wt	wt	wt	wt	wt
309	wt	wt	wt	wt	wt	wt	wt	wt	wt
312	wt	wt	wt	+	wt	wt	wt	wt	wt
315	wt	wt	wt	+	wt	+	wt	wt	wt
317	wt	wt	wt	wt	wt	wt	wt	wt	wt
318	wt	wt	wt	+	wt	+	wt	wt	wt
741	wt	wt	wt	wt	wt	wt	wt	wt	wt
744	wt	wt	wt	wt	wt	wt	wt	wt	wt
748	wt	wt	wt	wt	wt	wt	wt	wt	wt

Potential Applications

The results of these latest overexpression studies confirm our earlier conclusion that G481 and its equivalogs are excellent candidates for improvement of drought related stress tolerance in commercial species. Additionally, G481 related genes could also be used to manipulate flowering time.

G482 (SEQ ID NO: 89 and 90)

Published Information

G482 is a member of the HAP3 subgroup of the CCAAT-box binding transcription factor family, and is equivalent to AtHAP3b which was identified by Edwards et al. ((1998) *Plant Physiol.* 117:1015-1022) as an EST with homology to the yeast gene HAP3b. Edwards' northern blot data suggests that AtHAP3b is expressed primarily in roots. No other functional information regarding G482 is publicly available.

Experimental Observations

RT-PCR analysis of endogenous levels of G482 transcripts indicated that this gene was expressed constitutively in all tissues tested. A cDNA array experiment supported the RT-PCR derived tissue distribution data. G482 was not induced above basal levels in response to any environmental stress treatments tested.

We have now generated 35S lines for G482 using the two component system; two batches of T1 lines (321-341 and 341-360) were examined and many of the plants showed a striking acceleration of flowering (1-2 weeks sooner than wild-type) under 24 hour light conditions.

The early flowering effect was seen in 10/20 lines from the 321-341 set (#323, 325, 326, 327, 329, 330, 332, 333, 335, 336) and 7/20 lines from the 341-360 set (#341, 351, 355, 356, 357, 358, 360). Comparable effects on flowering time were also seen in each of three T2 populations (329, 330, 333) that were morphologically examined. In addition to the accelerated flowering, the majority of 35S::G482 lines also displayed a slight reduction in overall size; in fact a number of lines were very small (#321, 328, 331, 338, 339, 340 and #342, 344, 345, 349, 350) and did not survive to maturity.

All of the two-component lines showed segregation on selection plates in the T2 generation that was compatible with the transgene being present at a single locus.

G482 function was analyzed through its ectopic overexpression in plants under the control of a 35S promoter using the two component system.

In plate-based physiology assays, half of the two-component lines tested showed increased vigor on mannitol media and three lines performed better than wild-type in the heat germination assay.

It should be emphasized that we have observed stress-related tolerance phenotypes for several other G481 related genes including G485 and G1820. The similar effects seen when these genes are overexpressed strongly suggest a functional relationship between them.

Table 18. G482 35S, 2-components-supTfn

Line	Germ. in high NaCl	Germ. in high mannitol	Sucrose	ABA	Germ. in heat	Germ. in cold	Growth in heat	Drought	Chilling
324	wt	wt	wt	wt	wt	wt	wt	wt	wt
327	wt	wt	wt	wt	wt	wt	wt	wt	wt
330	wt	wt	wt	wt	wt	wt	wt	wt	wt
333	wt	wt	wt	wt	wt	wt	wt	wt	wt
343	wt	wt	wt	wt	wt	wt	wt	wt	wt
346	wt	+	wt	wt	wt	wt	+	wt	wt
347	wt	+	wt	wt	wt	wt	wt	wt	wt
351	wt	+	wt	wt	wt	wt	+	wt	wt
353	wt	+	wt	wt	wt	wt	wt	wt	wt
354	wt	+	wt	wt	wt	wt	+	wt	wt

Potential Applications

The results of this study bolster our conclusion that G481 and the related genes, including G482, are excellent candidates for improvement of drought related stress tolerance in commercial species.

Additionally, G482 could be useful for manipulating flowering time.

***Arabidopsis* G485 (SEQ ID NO: 2009 and 2010)**

G485 is paralog of G481, and is a member of the HAP3 subgroup of the CCAAT-box binding transcription factor family. It has been referenced as sequence 1042 from Patent Application WO0216655 on stress-regulated genes, transgenic plants and methods of use. G485 was reported therein to be cold responsive in a microarray analysis (Harper et al. (2002) Patent Application WO 0216655-A 1042 28-FEB-2002). No other functional information regarding G485 is publicly available.

During our earlier genomics program, we determined that plants overexpressing G485 had accelerated flowering, bolting up to 1 week earlier than wild-type or non-transformed plants grown under 24 hr lights. These studies, combined with related studies on plants lacking G485 expression have implicated G485 as both sufficient to act as a floral activator, and also necessary in that role within the plant.

Experimental Observations

The aim of this study was to re-assess the effects of overexpression of G485 using a two-component system and determine if this gene can confer enhanced stress tolerance in a manner comparable to G481.

We have now generated 35S lines for G485 using the two component system. These plants showed a comparable early flowering phenotype to that observed for 35S direct promoter fusion lines in our earlier genomics program.

Many of the 35S::G485 two-component lines exhibited a marked acceleration in the onset of flowering and generally formed flower buds 1-2 weeks sooner than wild type under continuous light

conditions. Many of the lines also showed a reduction in rosette biomass compared to wild type. In fact, three of twenty lines (308, 312, 320) showed a severe dwarf phenotype and did not survive to maturity. Early flowering was exhibited by 11/20 of the T1 lines (#301, 302, 303, 304, 306, 307, 309, 313, 315, 317, 319). The remaining lines appeared wild-type, apart from lines 310 and 314, which were noted to be slightly delayed in the onset of flowering. Line 14 was also infertile and failed to yield seed. Flowering time was also assessed in a number of T2 populations: plants from the T2-302, T2-305, T2-307, and T2-309 all displayed early flowering comparable to that seen in the parental lines. Plants from the T2-310 and T2-311 populations flowered at the same time as controls.

All of the ten two-component lines submitted for physiological assays showed segregation on selection plates in the T2 generation that was compatible with the transgene being present at a single locus.

In plate-based physiology assays, 35S::G485 two-component lines were more tolerant to salt stress in a germination assay compared to wild-type or non-transformed seedlings. Several salt tolerant lines were also less sensitive to sucrose, ABA, and cold stress in separate germination assays.

Table 19. G485 35S, 2-components-supTfn

Line	Germ. in high NaCl	Germ. in high mannitol	Sucrose	ABA	Germ. in heat	Germ. in cold	Growth in heat	Drought	Chilling
302	+	wt	wt	+	wt	wt	wt	wt	wt
305	+	wt	wt	+	wt	+	wt	wt	wt
307	+	wt	wt	wt	wt	wt	wt	wt	wt
309	wt	wt	wt	wt	wt	wt	wt	wt	wt
310	+	wt	wt	wt	wt	+	wt	wt	wt
311	+	wt	+	wt	wt	+	wt	wt	wt
316	+	wt	+	+	wt	+	wt	wt	wt
317	+	wt	wt	wt	wt	wt	wt	wt	wt
318	wt	wt	wt	+	wt	+	wt	wt	wt
319	+	wt	+	wt	wt	+	wt	wt	wt

Potential applications

Based on the results of these abiotic stress assays indicate that G481 and related genes, including G485, are excellent candidates for improvement of drought related stress tolerance in commercial species.

Additionally, G485 could be used to manipulate flowering time.

G489 (SEQ ID NO: 93 and 94)

Published Information

G489 was identified from a BAC sequence that showed high sequence homology to AtHAP5-like transcription factors in *Arabidopsis*. During our earlier genomics program, we observed that plants

overexpressing G489 were tolerant of NaCl and mannitol in separate germination assays. Morphologically, the plants were similar to wild-type or non-transformed plants.

Experimental Observations

Two sets of 35S::G489 lines (301-320 and 421-440) were generated. Lines 301-320 harbored the 35S direct promoter-fusion construct (P51). The second batch of plants (421-440) overexpressed G489 via the two component system. Neither of the two sets of 35S::G489 plants showed any consistent differences in morphology to wild type controls.

The function of G489 was analyzed through its ectopic overexpression in plants. G489 overexpressors were more tolerant to high NaCl stress, showing more root growth and leaf expansion compared with the controls in culture. Two well characterized ways in which NaCl toxicity is manifested in the plant is through general osmotic stress and potassium deficiency due to the inhibition of its transport. These G489 overexpressor lines were more tolerant to osmotic stress in general, showing more root growth on mannitol containing media. G489 overexpressors were also more tolerant to drought than wild-type control plants in soil-based assays.

RT-PCR analysis of endogenous levels of G489 transcripts indicated that this gene was expressed constitutively in all tissues tested. A cDNA array experiment confirmed the RT-PCR derived tissue distribution data. G489 was not induced above basal levels in response to stress treatments tested.

Several 35S::G489 lines derived from direct promoter fusions were created and characterized morphologically. These plants showed no consistent differences in morphology from wild type controls. As shown in Table 20, five of ten G489-overexpressing lines tested were better able to germinate in cold conditions. Three lines of more mature plants were also better able to tolerate cold conditions.

Table 20. G489 35S, Direct promoter-fusion and 2-components-supTfn

Line	Transfor-mation	Germ. in high NaCl	Germ. in high mannitol	Sucrose	ABA	Germ in heat	Germ in cold	Growth in heat	Drought	Chilling
301	Direct fusion	wt	wt	wt	wt	wt	wt	wt	wt	+
302	Direct fusion	wt	wt	wt	wt	wt	+	wt	wt	wt
303	Direct fusion	wt	wt	wt	wt	wt	+	wt	wt	wt
304	Direct fusion	wt	wt	wt	wt	wt	wt	wt	+	+
305	Direct fusion	wt	wt	wt	wt	wt	+	wt	+	+
308	Direct fusion	wt	wt	wt	wt	wt	+	wt	wt	wt
310	Direct fusion	wt	wt	wt	wt	wt	+	wt	wt	wt
314	Direct fusion	wt	wt	wt	wt	wt	wt	wt	wt	wt

316	Direct fusion	wt	wt	wt	wt	wt	wt	wt	wt	wt
317	Direct fusion	wt	wt	wt	wt	wt	wt	wt	wt	wt
421	2-comp	wt	wt	wt	wt	wt		wt	+	
422	2-comp	wt	wt	wt	wt	wt		wt	wt	
423	2-comp	wt	wt	wt	wt	wt		wt	wt	
424	2-comp	wt	wt	wt	wt	wt		wt	wt	
425	2-comp	wt	wt	wt	wt	wt		wt	wt	
426	2-comp	wt	wt	wt	wt	wt		wt	wt	
430	2-comp	wt	wt	wt	wt	wt		wt	wt	
434	2-comp	wt	wt	wt	wt	wt		wt	wt	
437	2-comp	wt	wt	wt	wt	wt		wt	+	
440	2-comp	wt	wt	wt	wt	wt	+	wt	wt	

Potential Applications

The results of this study bolster our conclusion that G481 and the related genes, including G489, are excellent candidates for improvement of drought related stress tolerance in commercial species.

G634 (SEQ ID NO: 127 and 128)

Published Information

G634 was initially identified as public partial cDNAs sequences for GTL1 and GTL2 which are splice variants of the same gene (Small et al (1998) *Proc. Natl. Acad. Sci.* U S A. 95:3318-3322). The published expression pattern of GTL1 shows that G634 is highly expressed in siliques and not expressed in leaves, stems, flowers or roots. There is no published information on the function of G634.

Closely Related Genes from Other Species

The closest non-*Arabidopsis* relative of G634 is the *O. sativa gt-2* gene (*EMBO J.* (1992) 11:4131-4144), which is proposed to bind and regulate the phyA promoter. In addition, the pea DNA-binding protein DF1 (13786451) shows strong homology to G634. The homology of these proteins to G634 extends to outside of the conserved domains and thus these genes are likely to be orthologs of G634.

Experimental Observations

The boundaries of G634 in were experimentally determined and the function of G634 was investigated by constitutively expressing G634 using the CaMV 35S promoter.

Three constructs were made for G634: P324, P1374 and P1717. P324 was found to encode a truncated protein. P1374 and P1717 represent full length splice variants of G634; P1374, the shorter of the two splice variants was used for the experiments described here. The longest available cDNA (P1717), confirmed by RACE, has the same ATG and stop codons as the genomic sequence.

Plants overexpressing G634 from construct P1374 showed a dramatic increase the density of trichomes, which additionally appear larger in size. The increase in trichome density was most noticeable on later arising rosette leaves, cauline leaves, inflorescence stems and sepals with the stem trichomes being more highly branched than controls. Approximately half of the primary transformants and two of three T2 lines showed the phenotype. Apart from slight smallness, there did not appear to be any other clear phenotype associated with the overexpression of G634. However, a reduction in germination was observed in T2 seeds grown in culture. It is not clear whether this defect was due to the quality of the seed lot tested or whether this characteristic is related to the transgene overexpression.

RT PCR data showed that G634 is potentially preferentially expressed in flowers and germinating seedlings, and induced by auxin. The role of auxin in trichome initiation and development has not been established in the published literature.

The increase in trichome density observed in G634 overexpressors suggested a possible role for this gene in drought-stress tolerance, a presumption subsequently confirmed in soil-based drought assays. We tested lines overexpressing G634 in a soil drought assay and found that they showed an enhanced performance versus wild type; G634 overexpressors recovered from the effects of a drought treatment significantly better than wild-type control plants. Additionally, our recent array experiments on plants undergoing a soil-drought experiment indicated that G634 shows a small but significant up-regulation specifically in the recovery phase following re-watering.

G634 overexpressing lines did not exhibit a shade avoidance phenotype when grown under light deficient in the red region of the visible spectrum. When the assay was repeated on individual lines, all three lines analyzed showed the phenotype and had short hypocotyls compared with wild-type seedlings. On control plates, seedlings from line 5 were small while lines 6 and 8 were comparable in size to wild-type seedlings.

Potential Applications

Trichome glands on the surface of many higher plants produce and secrete exudates that give protection from the elements and pests such as insects, microbes and herbivores. These exudates may physically immobilize insects and spores, may be insecticidal or ant-microbial or they may allergens or irritants to protect against herbivores. Trichomes have also been suggested to decrease transpiration by decreasing leaf surface air flow, and by exuding chemicals that protect the leaf from the sun.

Depending on the plant species, varying amounts of diverse secondary biochemicals (often lipophilic terpenes) are produced and exuded or volatilized by trichomes. These exotic secondary biochemicals, which are relatively easy to extract because they are on the surface of the leaf, have been widely used in such products as flavors and aromas, drugs, pesticides and cosmetics. One class of secondary metabolites, the diterpenes, can effect several biological systems such as tumor progression, prostaglandin synthesis and tissue inflammation. In addition, diterpenes can act as insect pheromones, termite allomones, and can exhibit neurotoxic, cytotoxic and antimitotic activities. As a result of this

functional diversity, diterpenes have been the target of research several pharmaceutical ventures. In most cases where the metabolic pathways are impossible to engineer, increasing trichome density or size on leaves may be the only way to increase plant productivity.

Thus, the use of G634 and its equivalogs to increase trichome density, size or type may therefore have profound utilities in so called *molecular farming* practices (i.e. the use of trichomes as a manufacturing system for complex secondary metabolites), and in producing resistant insect and herbivore resistant plants.

Of particular significance, G634 and its equivalogs may also be used to increase the osmotic stress, drought and shade tolerance of plants.

G682 (SEQ ID NO: 147 and 148)

Published Information

G682 was identified from the *Arabidopsis* BAC, AF007269, based on sequence similarity to other members of the Myb family within the conserved domain.

Experimental Observations

RT-PCR analysis of the endogenous levels of G682 transcripts indicated that this gene was expressed in all tissues tested, however, a very low level of transcript was detected in roots and shoots. Array tissue print data indicated that G682 was expressed primarily, but not exclusively, in flower tissue.

An array experiment was performed on one G682 overexpressing line. The data from this one experiment indicated that this gene could be a negative regulator of chloroplast development and/or light dependent development because the gene Albino3 and many chloroplast genes are repressed. Albino3 functions to regulate chloroplast development (Sundberg et al (1997) *Plant Cell* 9:717-730). The gene G682 was itself induced 20-fold. Other than a few additional transcription factors, very few genes are induced as a result of the ectopic expression of G682.

The function of G682 was analyzed through its ectopic overexpression in plants. G682 overexpressors were glabrous, had tufts of more root hairs and germinated better under heat stress conditions. Older plants were not more tolerant to heat stress compared to wild-type controls.

G682 overexpressors are glabrous, have tufts of more root hairs and germinated better under heat stress conditions. Older plants were not more tolerant to heat stress compared to wild-type controls. At the time these experiments were performed, it was suggested that further experiments were needed to address whether or not the heat germination phenotype of the G682 overexpressors was related to water deficit stress tolerance in the germinating seedling, and correlated with a possible drought tolerance phenotype. More recent experiments have shown that G682 overexpressors were more tolerant to water deprivation conditions in soil-based drought assays than wild-type plants, and two of three lines tested were significantly more drought tolerant than the wild-type controls.

In addition to the paralogous sequences disclosed above, orthologous sequences from other plant species were also identified using BLAST analysis. Such orthologous sequences, together with the paralogous sequences were determined to be members of the G682 TF family of Myb-related proteins (equivalogs). The paralogous sequences and the orthologous sequences were aligned using MACVECTOR software (Accelrys, Inc.). The software program also generated an exemplary consensus amino acid residue sequence of the aligned sequences.

As shown in Figures 3A and 3B, the orthologous sequences shared a consensus sequence with the conserved domain of G682 (amino acid residues 27-63 of SEQ ID NO:148) and also shared identity with regions flanking the conserved domain (flanking regions). In particular, G682 shared a region of the conserved domain with sequences from soy (*Glycine max*; SEQ ID NOs: 1084, 1085, 1086, 1083, 1087, and 1088), rice (*Oryza sativa*; SEQ ID NOs: 559, 1082, and 1081), and maize (corn) (*Zea mays*; SEQ ID NOs: 1089 and 1090).

An exemplary consensus of the conserved domain of the G682 TF family of Myb-related proteins is Val-Xaa-Met/Phe-Ser/Thr-Gln/Glu-Xaa-Glu-Glu-Asp-Leu-Val-Xaa-Arg-Met-His/Tyr-Lys/Arg-Leu-Val-Gly-Asp/Glu-Arg/Lys-Trp-Glu/Asp-Leu/Ile-Ile-Ala-Gly-Arg-Ile/Val-Pro-Gly-Arg, where Xaa is any amino acid residue. An alternative exemplary consensus of the conserved domain is Val-Xaa-Met/Phe-Ser/Thr-Gln/Glu-Xaa-Glu-Glu-Asp-Leu-Val-Ser-Arg-Met-His-Arg-Leu-Val-Gly-Asn-Arg-Trp-Glu-Leu-Ile-Ala-Gly-Arg-Ile-Xaa-Gly-Arg, where Xaa is any amino acid residue. A further alternative exemplary consensus of the conserved domain is Val-Xaa-Met/Phe-Ser/Thr-Gln/Glu-Xaa-Glu-Glu-Asp-Leu-Val-Ser-Arg-Met-Tyr-Xaa-Leu-Val-Gly-Asn/Glu-Arg-Trp-Ser-Leu-Ile-Ala-Gly-Arg-Ile-Pro-Gly-Arg, where Xaa is any amino acid residue.

Potential Applications

The potential utility of this gene or its equivalogs is to confer heat tolerance to germinating seeds. G682 or its equivalogs could be used to alter trichome number and distribution in plants. Trichome glands on the surface of many higher plants produce and secrete exudates, which give protection from the elements and pests such as insects, microbes and herbivores. These exudates may physically immobilize insects and spores, may be insecticidal or ant-microbial or they may allergens or irritants to protect against herbivores. Trichomes have also been suggested to decrease transpiration by decreasing leaf surface air flow, and by exuding chemicals that protect the leaf from the sun.

G864 (SEQ ID NO: 167 and 168)

Published Information

G864 was identified in an *Arabidopsis* EST (H37693). G864 appears as gene AT4g23750 in the annotated sequence of *Arabidopsis* chromosome 4 (AL161560).

Experimental Observations

G864 was discovered and initially identified as a public *Arabidopsis* EST. G864 was ubiquitously expressed, and was not significantly induced under any of the conditions tested.

The complete sequence of G864 was determined, and G864 was found to be related to two additional *Arabidopsis* AP2/EREBP genes, G1421 and G1755. The function of G864 was analyzed using transgenic plants in which this gene was expressed under the control of the 35S promoter. G864 overexpressing plants exhibited a variety of phenotypic alterations. They were smaller than wild-type plants, and those with the strongest phenotypes were classified as dwarf. However, G864 overexpressing lines showed more seedling vigor in a heat stress tolerance germination assay compared to wild-type controls. Conversely, G864 overexpressing lines were also somewhat more sensitive to chilling. One of the three T2 lines analyzed showed significant increase in fucose and arabinose levels in leaves.

In soil-based assays, G864 overexpressing plants were significantly more drought tolerant than wild-type control plants.

Potential Applications

The germination of many crops is very sensitive to temperature. A gene that would enhance germination in hot conditions such as G864 or its equivalogs would be useful for crops that are planted late in the season or in hot climates. G864 and its equivalogs may also be used to improve the drought or other osmotic stress tolerance of plants.

G867 (SEQ ID NO: 169 and 170)

Published Information

G867 corresponds to RAV1 (Kagaya et al. (1999) *Nucleic Acids Res.* 27: 470-478). G867/RAV1 belongs to a small subgroup within the AP2/EREBP family of transcription factors, whose distinguishing characteristic is that its members contain a second DNA-binding domain, in addition to the conserved AP2 domain, that is related to the B3 domain of VP1/ABI3 (Kagaya et al. (1999) *supra*). It has been shown that the two DNA-binding domains of RAV1 can separately recognize each of two motifs that constitute a bipartite binding sequence and together cooperatively enhance its DNA-binding affinity and specificity (Kagaya et al. (1999) *supra*).

Experimental Observations

G867 was discovered and initially identified as a public *Arabidopsis* EST. G867 appeared to be constitutively expressed at medium levels.

G867 was first characterized using a line that contained a T-DNA insertion in the gene. The insertion in that line resided immediately downstream of the conserved AP2 domain, and would therefore be expected to result in a severe or null mutation. G867 knockout mutant plants did not show significant

changes in overall plant morphology, significant differences between these plants and control plants have not been detected in any of the assays that have been performed so far.

Subsequently, the function of G867 was analyzed using transgenic plants in which this gene was expressed under the control of the 35S promoter. G867 overexpressing lines were morphologically wild-type and no phenotypic alterations in G867 overexpressing lines were detected in the biochemical assays that were performed. However, G867 overexpressing lines showed increased seedling vigor (manifested by increased expansion of the cotyledons) in germination assays on both high salt and high sucrose containing media, compared to wild-type controls.

The *Arabidopsis* paralogs G1930 (SEQ ID NO: 369) and G9 (SEQ ID NO: 1949) also showed stress related phenotypes. G9 exhibited increased root biomass, and thus could be used to produce better plant growth under adverse osmotic conditions. Genetic and physiological evidence indicates that roots subjected to various stresses, including water deficit, alter the export of specific compounds, such as ACC and ABA, to the shoot, via the xylem Bradford et al. (1980) *Plant Physiol.* 65: 322-326; Schurr et al. (1992) *Plant Cell Environ.* 15, 561-567).

G1930 plants responded to high NaCl and high sucrose on plates with more seedling vigor, and root biomass compared to wild-type control plants; this phenotype was identical to that seen in 35S::G867 lines. These results indicate a general involvement of this clade in abiotic stress responses:

The polypeptide sequences of G1930 and G9 share 72% (249/345 residues) and 64% (233/364 residues) with G867, respectively. The conserved domains of G1930 and G9 are 86% (56/65 residues) and 86% (56/65 residues) identical with the conserved domain of G867, respectively.

In addition to the paralogous sequences disclosed above, orthologous sequences from other plant species were also identified using BLAST analysis. Such orthologous sequences, together with the paralogous sequences were determined to be members of the G867 TF family of AP2 proteins (equivalogs). The paralogous sequences and the orthologous sequences were aligned using MACVECTOR software (Accelrys, Inc.). The software program also generated an exemplary consensus amino acid residue sequence of the aligned sequences.

As shown in Figures 4A, 4B, 4C, and 4D, the orthologous sequences shared a consensus sequence with the conserved domain of G867 (amino acid residues 59-116 of SEQ ID NO:170) and also shared identity with regions flanking the conserved domain (flanking regions). In particular, G867 shared a region of the conserved domain with sequences from soy (*Glycine max*; SEQ ID NOs: 1184, 1183, and 1182), rice (*Oryza sativa*; SEQ ID NOs: 1176, 1177, and 1178), and maize (corn) (*Zea mays*; SEQ ID NOs: 1186 and 1185).

An exemplary consensus of the conserved domain of the G867 TF family of AP2 proteins is Ser-Ser-Lys/Arg-Tyr/Phe-Gly-Val-Val-Pro-Gln-Pro-Asn-Gly-Arg-Typ-Gly-Ala-Gln-Ile-Tyr-Glu-Lys/Arg-His-Gln-Arg-Val-Trp-Leu-Gly-Thr-Phe-Xaa-Glu/Asp-Glu-Glu/Asp-Glu/Asp-Ala-Ala/Val-Arg-Ala/Ser-Tyr-Asp-Val/Ile-Ala/Val-Val/Ala-Xaa-Arg-Phe/Tyr-Arg-Arg/Gly-Arg-Asp-Ala-Val-Thr/Val-Asn-Phe-Lys/Arg, where Xaa is any amino acid residue. An alternative exemplary consensus of the conserved

domain is Ser-Ser-Lys/Arg-Tyr/Phe-Gly-Val-Val-Pro-Gln-Pro-Asn-Gly-Arg-Typ-Gly-Ala-Gln-Ile-Tyr-Glu-Lys/Arg-His-Gln-Arg-Val-Trp-Leu-Gly-Thr-Phe-Xaa-Glu/Asp-Glu-Glu/Asp-Ala-Ala-Ala-Arg-Ala-Tyr-Asp-Val/Ile-Ala/Val-Val/Ala-Xaa-Arg-Phe/Tyr-Arg-Arg/Gly-Arg-Asp-Ala-Val-Thr/Val-Asn, where Xaa is any amino acid residue. A further alternative exemplary consensus of the conserved domain is Ser-Ser-Lys/Arg-Tyr/Phe-Gly-Val-Val-Pro-Gln-Pro-Asn-Gly-Arg-Typ-Gly-Ala-Gln-Ile-Tyr-Glu-Lys/Arg-His-Gln-Arg-Val-Trp-Leu-Gly-Thr-Phe-Xaa-Gly-Glu-Ala/Asp-Glu/Asp-Ala-Ala/Val-Arg-Ala-Tyr-Asp-Val-Ala-Ala-Gln-Arg-Phe/Tyr-Arg-Arg/Gly-Arg-Asp-Ala-Val-Thr/Val-Asn-Phe-Arg, where Xaa is any amino acid residue.

Potential Applications

G867 or its equivalogs could be used to increase or facilitate seed germination and seedling growth under adverse environmental conditions, in particular salt stress.

G867 or its equivalogs may also be used to modify sugar sensing.

G912 (SEQ ID NO: 185 and 186)

Published Information

G912 was identified in the sequence of P1 clone MSG15 (GenBank accession number AB015478; gene MSG15.6).

Experimental Observations

G912 was recognized as the AP2/EREBP gene most closely related to *Arabidopsis* CBF1, CBF2, and CBF3 (Stockinger et al (1997) *Proc. Natl. Acad. Sci. USA* 94:1035-1040; Gilmour et al. (1998) *Plant J.*16:433-442). In fact, G912 is the only other AP2/EREBP transcription factor for which sequence similarity with CBF1, CBF 2, and CBF3 extends beyond the conserved AP2 domain.

The function of G912 was studied using transgenic plants in which this gene was expressed under the control of the 35S promoter. Plants overexpressing G912 were more freezing and drought tolerant than the wild-type controls, but were also small, dark green, and late flowering. There was a positive correlation between the degree of growth impairment and the freezing tolerance. In addition, G912 expression appeared to be induced by cold, drought, and osmotic stress.

In addition, G912 overexpressing plants also exhibited a sugar sensing phenotype: reduced seedling vigor and cotyledon expansion upon germination on high glucose media.

These results mirror the extensive body of work that has shown that CBF1, CBF2, and CBF3 are involved in the control of the low-temperature response in *Arabidopsis*, and that those genes can be used to improve freezing, drought, and salt tolerance in plants (Stockinger et al., (1997) *Proc. Natl. Acad. Sci. USA* 94:1035-1040; Gilmour et al. (1998) *Plant J.*16:433-442; Jaglo-Ottosen et al. (1998) *Science.* 280:104-106; Liu et al. (1998) *Plant Cell.* 10:1391-1406, Kasuga et al. (1999) *Nat. Biotechnol.* 17:287-291).

The polypeptide sequences of G40, G41, and G42 share 71% (140 of 195 residues), 68% (144 of 211 residues), and 65% (147 of 224 residues) identity with G912, respectively. The conserved domains of G40, G41, and G42 share 94% (64 of 68 residues), 92% (63 of 68 residues), and 94% (64 of 68 residues) identity with G912, respectively.

In addition to the paralogous sequences disclosed above, orthologous sequences from other plant species were also identified using BLAST analysis. Such orthologous sequences, together with the paralogous sequences were determined to be members of the G912 TF family of AP2/EREBP proteins (equivalogs). The paralogous sequences and the orthologous sequences were aligned using MACVECTOR software (Accelrys, Inc.). The software program also generated an exemplary consensus amino acid residue sequence of the aligned sequences.

As shown in Figures 5A, 5B, 5C, 5D, 5E, and 5F, the orthologous sequences shared a consensus sequence with the conserved domain of G912 (amino acid residues 51-118 of SEQ ID NO: 186) and also shared identity with regions flanking the conserved domain (flanking regions). In particular, G912 shared a region of the conserved domain with sequences from soy (*Glycine max*; SEQ ID NOs: 1238, 1242, 1240, 1241, and 1243), rice (*Oryza sativa*; SEQ ID NOs: 1222, 1223, 1232, 1221, 1231, 1227, 1235, 1230, 1229, and 1228), and maize (corn) (*Zea mays*; SEQ ID NOs: 1246, 1247, 1244, and 1245).

An exemplary consensus of the conserved domain of the G912 TF family of AP2/EREBP proteins is His-Pro-Ile/Val-Tyr/Phe-Arg/Lys-Gly-Val-Arg-Gln/Arg-Arg-Gly/Asn-Xaa$_{(1-3)}$-Lys/Arg-Trp-Val-Cys/Ser-Glu-Val/Leu-Arg-Glu/Val-Pro-Asn-Lys-Xaa$_{(2)}$-Arg-Ile/Leu-Trp-Leu-Gly-Thr-Phe/Tyr-Xaa$_{(2)}$-Ala/Pro-Glu-Met-Ala-Ala-Arg-Ala-His-Asp-Val-Ala-Ala/Met-Leu/Met-Ala-Leu-Arg-Gly-Xaa$_{(1-8)}$-Ala-Cys-Leu-Asn-Phe-Ala-Asp-Ser-Xaa$_{(1-5)}$-Val/Ile-Pro/Asp, where Xaa is any amino acid residue. An alternative exemplary consensus of the conserved domain is His-Pro-Ile/Val-Tyr/Phe-Arg/Lys-Gly-Val-Arg-Xaa-Arg-Gly/Asn-Xaa$_{(1-3)}$-Lys/Arg-Trp-Val-Cys/Ser-Glu-Val/Leu-Arg-Glu/Val-Pro-Xaa$_{(1-5)}$-Arg-Ile/Leu/Phe-Trp-Leu-Gly-Thr-Phe/Tyr-Xaa$_{(2)}$-Ala/Pro-Glu-Xaa-Ala-Ala-Arg-Ala-His-Asp-Val-Ala-Ala/Met-Leu/Met-Ala-Leu-Arg-Gly-Xaa$_{(1-8)}$-Ala-Cys/Ser-Leu-Asn-Phe-Ala-Asp-Ser-Xaa$_{(1-5)}$-Val/Ile-Pro/Asp, where Xaa is any amino acid residue.

An exemplary flanking region consensus sequence of the G912 TF family of AP2/EREBP proteins is Pro-Lys-Xaa-Xaa-Ala-Gly-Arg (amino acids 37-43 of SEQ ID NO: 186), or Ala-Gly-Arg-Xaa-Lys-Phe (amino acids 41-46 of SEQ ID NO: 186) or Glu-Thr-Arg-His-Pro (amino acids 48-52 of SEQ ID NO: 186), where Xaa is any amino acid residue.

Potential Applications

G912 or its equivalogs could be used to improve plant tolerance to cold, freezing, drought, and salt stress. In addition, G912 or its equivalogs could be used to change a plant's flowering time and size.

G913 (SEQ ID NO: 187 and 188)

Published Information

G913 was identified in the sequence of clone MSG15; it corresponds to gene MSG15.10 (GenBank PID BAB11050).

Experimental Observations

The cDNA sequence of G913 was determined. To investigate the function(s) of G913, this gene was expressed under the control of the 35S promoter in transgenic plants. G913 overexpressing plants had dark green leaves that occasionally curled downward. These plants showed a delay in flowering and were also late senescing. Overexpressing G913 lines were more freezing tolerant and more drought tolerant than the wild-type controls.

In an ethylene sensitivity assay where the plants were tested for a triple response phenotype on plates containing ACC, G913 overexpressing plants showed stunting and curling in the hypocotyl region that was more exaggerated than the wild type triple response.

Potential Applications

G913 or its equivalogs could be used to improve plant tolerance to freezing and drought. G913 could also be used to manipulate the ethylene response.

G913 or its equivalogs may be used to delay flowering or senescence in plants. Extending vegetative development could bring about large increases in yields.

Additionally, a major concern is the escape of transgenic pollen from GMOs to wild species or so-called organic crops. Systems that prevent vegetative transgenic crops from flowering would eliminate this worry.

G922 (SEQ ID NO: 189 and 190)

Published Information

G922 corresponds to Scarecrow-like 3 (SCL3) first described by Pysh et al. (GenBank accession number AF036301; (1999) *Plant J.* 18: 111-119). Northern blot analysis results show that G922 is expressed in siliques, roots, and to a lesser extent in shoot tissue from 14 day old seedlings. Pysh et al did not test any other tissues for G922 expression. In situ hybridization results showed that G922 was expressed predominantly in the endodermis in the root tissue. This pattern of expression was very similar to that of SCARECROW (SCR), G306. Experimental evidence indicated that the co-localization of the expression is not due to cross-hybridization of the G922 probe with G306. Pysh et al proposed that G922 may play a role in epidermal cell specification and that G922 may either regulate or be regulated by G306.

The sequence for G922 can also be found in the annotated BAC clone F11F12 from chromosome 1 (GenBank accession number AC012561). The sequence for F11F12 was submitted to GenBank by the DNA Sequencing and Technology Center at Stanford University.

Experimental Observations

The function of this gene was analyzed using transgenic plants in which G922 was expressed under the control of the 35S promoter. Transgenic plants overexpressing G922 were more salt tolerant than wild-type plants as determined by a root growth assay on MS media supplemented with 150 mM NaCl. Plant overexpressing G922 also were more tolerant to osmotic stress as determined by germination assays in salt-containing (150 mM NaCl) and sucrose-containing (9.4%) media. Morphologically, plants overexpressing G922 had altered leaf morphology, coloration, fertility, and overall plant size. In wild-type plants, expression of G922 was induced by auxin, ABA, heat, and drought treatments. In non-induced wild-type plants, G922 was expressed constitutively at low levels.

The high salt assays suggested that this gene would confer drought tolerance, a supposition confirmed by soil-based assays, in which G922-overexpressing plants were significantly healthier after water deprivation treatment than wild-type control plants.

Potential Applications

Based upon results observed in plants overexpressing G922 or its equivalogs could be used to alter salt tolerance, tolerance to osmotic and drought stress, and leaf morphology in other plant species.

G926 (SEQ ID NO: 191 and 192)

Published Information

G926 is equivalent to Hap2a (Y13720), a member of the CCAAT-box binding transcription factor family. The gene was identified by Edwards et al. ((1998) *Plant Physiol.* 117: 1015-1022), who demonstrated that G926 or AtHap2a were able to functionally complement a Hap2 deficient mutant of yeast suggesting that there is functional conservation between these proteins from diverse organisms. In addition, the AtHap2a gene was shown to be ubiquitously expressed in *Arabidopsis*.

Experimental Observations

Consistent with the published expression pattern (Edwards et al. (1998) *Plant Physiol.* 117: 1015-1022), G926 was determined to be ubiquitously expressed and transcript levels appeared to be unaltered by any environmental stress-related condition tested. A line homozygous for a T-DNA insertion in G926 was used to determine the function of this gene.

The G926 knockout mutant line was morphologically wild-type. Physiological analysis revealed that in the presumed absence of G926 function, the plants became more tolerant to high osmotic conditions during germination. This osmotic stress tolerance could be related to the plant's apparent

insensitivity to the growth hormone ABA. This was the second instance where a member of a CCAAT-box protein complex altered the plants osmotic stress response and ABA sensitivity during germination.

ABA plays an important regulatory role in the initiation and maintenance of seed dormancy. Lopez-Molina, L. et al. ((2001) *Proc. Natl. Acad. Sci. USA* 98: 4782-4787) describe a bZIP transcription factor, ABI5, that is involved in maintaining seeds in a quiescent state, preventing germination under adverse conditions such as drought stress. It is possible G926 also functions as part of this checkpoint for the germinating seeds and loss of G926 function promotes germination regardless of the osmotic status of the environment.

Potential Applications

G926 or its equivalogs could be used to improve plant tolerance to drought, and salt stress.

Evaporation from the soil surface causes upward water movement and salt accumulation in the upper soil layer where the seeds are placed. Thus, germination normally takes place at a salt concentration much higher than the mean salt concentration of in the whole soil profile. Increased salt tolerance during the germination stage of a crop plant would impact survivability and yield.

G975 (SEQ ID NO: 199 and 200)

Published Information

G975 has appeared in the sequences released by the *Arabidopsis* Genome Initiative (BAC F9L1, GenBank accession number AC007591).

Experimental Observations

G975 was expressed in flowers and, at lower levels, in shoots, leaves, and siliques. GC-FID and GC-MS analyses of leaves from G975 overexpressing plants showed that the levels of C29, C31, and C33 alkanes were substantially increased (up to ten-fold) compared with control plants. A number of additional compounds of similar molecular weight, presumably also wax components, also accumulated to significantly higher levels in G975 overexpressing plants. C29 alkanes constituted close to 50% of the wax content in wild-type plants (Millar et al. (1998) *Plant Cell* 11:1889-1902), suggesting that a major increase in total wax content occurred in the G975 transgenic plants. However, the transgenic plants had an almost normal phenotype (although small morphological differences were detected in leaf appearance), indicating that overexpression of G975 was not deleterious to the plant. Overexpression of G975 did not cause the dramatic alterations in plant morphology that had been reported for *Arabidopsis* plants in which the FATTY ACID ELONGATION1 gene was overexpressed (Millar et al. 1998, *Plant Cell* 11:1889-1902). G975 may regulate the expression of some of the genes involved in wax metabolism. One *Arabidopsis* AP2 sequence (G1387) that is significantly more closely related to G975 than the rest of the members of the AP2/EREBP family is predicted to have a function and a use related to that of G975.

G975 overexpressing plants were significantly more drought tolerant than wild-type control plants in soil-based drought assays.

Potential Applications

G975 or its equivalogs can be used to manipulate wax composition, amount, or distribution, which in turn can modify plant tolerance to drought and/or low humidity or resistance to insects, as well as plant appearance (shiny leaves).

G975 or its equivalogs can also be used to specifically alter wax composition, amount, or distribution in those plants and crops from which wax is a valuable product.

G993 (SEQ ID NO: 2071 and 2072)

G993 is a putative paralog of G867. No genetic analysis of the locus has been published.

During our earlier genomics program, we observed that G993 overexpression lines exhibited a number of morphological abnormalities. The aim of this study was to re-assess 35S::G993 transformants (using a greater number of lines), and determine whether overexpression of the gene could confer enhanced stress tolerance in a comparable manner to G867.

New overexpression lines have been obtained using a direct promoter fusion construct. These lines exhibited similar phenotypes to those observed during our phase I genomics studies and were generally small, slow developing, and poorly fertile. Some lines also showed features that suggested a disruption in light regulated development, such as long hypocotyls and alterations in leaf orientation.

We have tested ten of the 35S::G993 direct-fusion lines under a variety of plate based treatments. Eight of these lines out-performed controls in one or more plate based stress assays including germination on salt, germination on sucrose, germination in the cold, or growth under chilling conditions. Interestingly, two 35S::G993 lines showed a very dramatic increase in root hair density when grown on regular MS plates (this phenotype was comparable to that observed in 35S::G9 lines during our phase I genomics program). Interestingly, though, the two lines that exhibited increased root hair development did not show a positive result in the stress assays.

It should be emphasized that we have obtained comparable developmental effects as well as a strong enhancement of drought related stress tolerance in all four of the *Arabidopsis* genes in the G867 study group (G9, G867, G993, and G1930). The almost identical phenotypic effects produced by these genes strongly suggest that they are functionally equivalent.

Table 21. G993 35S, Direct promoter-fusion

Line	Germ. in high NaCl	Germ. in high mannitol	Sucrose	ABA	Germ. in heat	Germ. in cold	Growth in heat	Drought	Chilling
321	+	wt	wt	wt	wt	wt	wt	wt	+
322	wt	wt	+	wt	wt	+	wt	wt	+
324	wt	wt	wt	wt	wt	wt	wt	wt	+

326	+	wt	wt	wt	wt	wt	wt	wt	+
330	wt	wt	wt	wt	wt	wt	wt	wt	wt
331	+	wt	+	wt	wt	wt	wt	wt	+
334	+	wt	++	wt	wt	++	wt	wt	wt
335	+	wt	++	wt	wt	+	wt	wt	wt
337	wt	wt	wt	wt	wt	wt	wt	wt	wt
340	+	wt	+	wt	wt	wt	wt	wt	+

Potential applications

Based on the results of our overexpression studies, G867 and related sequences such as G993 are excellent candidates for improvement of drought and cold-related stress tolerance in commercial species. The morphological effects associated with their overexpression suggests that tissue-specific or conditional promoters might be used to optimize the utility of these genes.

The increased root hair production seen in 35S::G993 lines indicates that the gene might be used to enhance root growth and differentiation and might thereby improve performance under other stresses, such as low nutrient availability.

G1048 (SEQ ID NO: 2515 and 2516)

Published Information

G1048 (AT1G42990) was initially identified as public partial EST T88194 and in BAC F13A11 (GenBank accession AC068324) released by the *Arabidopsis* Genome Initiative. There is no published information on the function(s) of G1048.

Experimental Observations.

RT-PCR expression analysis indicated that G1048 was constitutively expressed and not induced by any condition tested. At that time, the function of G1048 was investigated by constitutively expressing G1048 using the 35S promoter. Plants overexpressing G1048 were not significantly different to controls in any assay performed.

In initial experiments, G1048 overexpressing lines did not exhibit a shade avoidance phenotype when grown under light deficient in the red region of the visible spectrum . This effect was seen in two repeat experiments on a batch of mixed seed from three independent lines.

In repeat experiments, 35S::G1048 lines grown under white light versus light deficient in red light were analyzed. A significant shade tolerance phenotype was observed, indicating that G1048 might be involved in the transcriptional regulation of response to shade or light quality. G1048 is potentially related to *HY5* (Oyama et al. (1997) *Genes Dev.* 11:2983-2995), a gene that is well established to be involved in light regulated development.

G1069 (SEQ ID NO: 221 and 222)

Published Information

The sequence of G1069 was obtained from EU *Arabidopsis* sequencing project, GenBank accession number Z97336, based on its sequence similarity within the conserved domain to other AT-Hook related proteins in *Arabidopsis*.

Experimental Observations

The sequence of G1069 was experimentally determined and the function of G1069 was analyzed using transgenic plants in which G1069 was expressed under the control of the 35S promoter.

Plants overexpressing G1069 showed changes in leaf architecture, reduced overall plant size, and retarded progression through the life cycle. This is a common phenomenon for most transgenic plants in which AT-HOOK proteins are overexpressed if the gene is predominantly expressed in root in the wild-type background. G1069 was predominantly expressed in roots, based on analysis of RT-PCR results. To minimize these detrimental effects, G1069 may be overexpressed under a tissue specific promoter such as root- or leaf-specific promoter or under inducible promoter.

One of G1069 overexpressing lines showed more tolerance to osmotic stress when they were germinated in high sucrose plates. This line also showed insensitivity to ABA in a germination assay.

Potential Applications

The osmotic stress results indicate that G1069 could be used to alter a plant's response to water deficit conditions and, therefore, the gene or its equivalogs could be used to engineer plants with enhanced tolerance to drought, salt stress, and freezing.

G1069 affects ABA sensitivity, and thus when transformed into a plant the gene or its equivalogs may diminish cold, drought, oxidative and other stress sensitivities, and also be used to alter plant architecture, and yield.

G1073 (SEQ ID NO: 223 and 224)

Published Information

G1073 has been identified in the sequence of a BAC clone from chromosome 4 (BAC clone F23E12, gene F23E12.50, GenBank accession number AL022604), released by EU *Arabidopsis* Sequencing Project.

Experimental Observations

The function of G1073 was analyzed using transgenic plants in which G1073 was expressed under the control of the 35S promoter. Transgenic plants overexpressing G1073 were substantially larger than wild-type controls, with at least a 60% increase in biomass. The increased mass of 35S::G1073 transgenic plants was attributed to enlargement of multiple organ types including leaves, stems, roots and

floral organs. Petal size in the 35S::G1073 lines was increased by 40-50% compared to wild type controls. Petal epidermal cells in those same lines were approximately 25-30% larger than those of the control plants. Furthermore, 15-20% more epidermal cells per petal were produced compared to wild type. Thus, at least in petals, the increase in size was associated with an increase in cell size as well as in cell number. Additionally, images from the stem cross-sections of 35S::G1073 plants revealed that cortical cells are large and that vascular bundles contained more cells in the phloem and xylem relative to wild type

Seed yield was increased compared to control plants. 5S::G1073 lines showed an increase of at least 70% in seed yield. This increased seed production was associated with an increased number of siliques per plant, rather than seeds per silique.

Flowering of G1073 overexpressing plants was delayed. Leaves of G1073 overexpressing plants were generally more serrated than those of wild-type plants. Improved drought tolerance was observed in 35S::G1073 transgenic lines.

Potential Applications

Transgenic plants overexpressing G1073 are large and late flowering with serrated leaves. Large size and late flowering produced as a result of G1073 or equivalog overexpression would be extremely useful in crops where the vegetative portion of the plant is the marketable portion (often vegetative growth stops when plants make the transition to flowering). In this case, it would be advantageous to prevent or delay flowering with the use of this gene or its equivalogs in order to increase yield (biomass). Prevention of flowering by this gene or its equivalogs would be useful in these same crops in order to prevent the spread of transgenic pollen and/or to prevent seed set. This gene or its equivalogs could also be used to manipulate leaf shape and drought tolerance.

G1075 (SEQ ID NO: 225 and 226)

Published Information

The sequence of G1075 was obtained from the *Arabidopsis* genome sequencing project, GenBank accession number AC004667, based on its sequence similarity within the conserved domain to other AT-Hook related proteins in *Arabidopsis*.

Experimental Observations

The function of G1075 was analyzed using transgenic plants in which G1075 was expressed under the control of the 35S promoter. Overexpression of G1075 produced very small, sterile plants. Pointed leaves were noted in some seedlings, and twisted or curled leaves and abnormal leaf serrations were noted in rosette stage plants. Bolts were short and thin with short internodes. Flowers from severely affected plants had reduced or absent petals and stamen filaments that partially or completely fail to

elongate. Because of the severe phenotypes of these T1 plants, no T2 seed was produced for physiological and biochemical analysis.

RT-PCR analysis indicated that G1075 transcripts are found primarily in roots. The expression of G1075 appeared to be induced by cold and heat stresses.

Potential Applications

G1075 or its equivalogs could be used to modify plant architecture and development, including flower structure. If expressed under a flower-specific promoter, the gene or its equivalogs might also be useful for engineering male sterility. Because expression of G1075 is root specific, its promoter could be useful for targeted gene expression in this tissue.

***Arabidopsis* G1364 (SEQ ID NO: 2101 and 2102)**

G1364, a putative paralog of G481, is a member of the HAP3 subgroup of the CCAAT-box binding transcription factor family.

Experimental Observations

The aim of the present study was to reevaluate the effects of G1364 overexpression and determine whether the gene confers similar effects to G481. We have now obtained two sets of 35S::G1364 lines using a two-component approach. A significant number of these transformants showed delayed flowering, indicating that G1364 can act as a repressor of the floral transition. Plants from this set also senesced later than wild type. In other respects, these transformants showed wild-type morphology.

As shown in Table 22, two G1364-overexpressing lines tested were less sensitive to ABA than wild-type or non-transformed plants.

Table 22. G1364 35S, 2-components-supTfn

Line	Germ. in high NaCl	Germ. in high mannitol	Sucrose	ABA	Germ. in heat	Germ. in cold	Growth in heat	Drought	Chilling
341	wt	wt	wt	wt	wt	wt	wt	wt	wt
342	wt	wt	wt	wt	wt	wt	wt	wt	wt
343	wt	wt	wt	wt	wt	wt	wt	wt	wt
344	wt	wt	wt	wt	wt	wt	wt	wt	wt
345	wt	wt	wt	wt	wt	wt	wt	wt	wt
346	wt	wt	wt	wt	wt	wt	wt	wt	wt
423	wt	wt	wt	wt	wt	wt	wt	wt	wt
431	wt	wt	wt	wt	wt	wt	wt	wt	wt
432	wt	wt	wt	+	wt	wt	wt	wt	wt
435	wt	wt	wt	+	wt	wt	wt	wt	wt

Potential Applications

Based on the results of the ABA germination assay, G481 and related sequences such as G1364 may be used to improve the stress tolerance in plants.

G1364 might also be used to modify flowering time related traits.

G1411 (SEQ ID NO: 269 and 270)

Published Information

G1411 was identified in the sequence of TAC clone K22G18 (GenBank accession number AB022212).

Experimental Observations

The complete sequence of G1411 was determined. The function of G1411 was analyzed using transgenic plants in which this gene was expressed under the control of the 35S promoter. G1411 overexpressing plants were smaller than wild-type controls and showed reduced apical dominance: axillary shoots develop prematurely amongst primary rosette leaves, resulting in a bushy plant. G1411 overexpressing plants behaved like the corresponding wild-type controls in all physiological and biochemical assays that were performed.

Potential Applications

G1411 or its equivalogs could be used to manipulate plant architecture.

G1451 (SEQ ID NO: 277 and 278)

Published Information

G1451 is ARF8, a member of the ARF class of proteins with a VP1-like N-terminal domain and a C-terminal domain with homology to Aux/IAA proteins. ARF8, like several other ARFs, contains a glutamine-rich central domain that can function as a transcriptional activation domain (1). ARF8 was shown to bind to an auxin response element (2). It was also shown that a truncated version of ARF8 lacking the DNA binding domain but containing the activation domain and the C-terminal domain could activate transcription on an auxin responsive promoter, presumably through interactions with another factor bound to the auxin response element (1). ARF8 is closely related in sequence to ARF6 (2).

Experimental Observations

G1451 was expressed throughout the plant, with the highest expression in flowers. Transcripts of G1451 were induced in leaves by a variety of stress conditions. A line homozygous for a T-DNA insertion in G1451 was used to determine the function of this gene. The T-DNA insertion of G1451 is approximately one-fifth of the way into the coding sequence of the gene and therefore is likely to result in a null mutation.

As measured by NIR, G1451 knockout mutants had increased total combined seed oil and seed protein content compared to wild-type plants.

Potential Applications

G1451 or its equivalogs may be used to alter seed oil and protein content, which may be very important for the nutritional value and production of various food products

G1451 or its equivalogs could also be used to increase plant biomass. Large size is useful in crops where the vegetative portion of the plant is the marketable portion since vegetative growth often stops when plants make the transition to flowering.

G1543 (SEQ ID NO: 303 and 304)

Published Information

G1543 was identified as a novel homeobox gene within section 3 of 255 from the complete sequence of Chromosome II (GenBank accession number AC005560, released by the *Arabidopsis* Genome Initiative).

Experimental Observations

The ends of G1543 were determined by RACE and a full-length cDNA was isolated by PCR from mixed cDNA. The encoded 275 amino acid product was found to be a member the HD-ZIP class II group of HD proteins. The public annotation for this gene was incorrect; the protein predicted in the BAC report was only 162 amino acids in length.

RT-PCR analysis revealed that G1543 was expressed ubiquitously but was up-regulated in response to auxin applications.

The function of G1543 was analyzed using transgenic plants in which the gene was expressed under the control of the 35S promoter. 35S::G1543 *Arabidopsis* plants exhibited a range of phenotypes; most consistently, however, the plants possessed dark green leaves and an altered branching pattern that led to a shorter more compact stature. These morphological phenotypes, along with the expression data, implicate G1543 as a component of a growth or developmental response to auxin.

Biochemical assays reflected the changes in leaf color noted during morphological analysis. All three T2 lines examined displayed increased levels of leaf chlorophylls and carotenoids. Additionally, one of three lines had a decrease in seed oil combined with an increase in seed protein. A repeat experiment verified the altered seed oil and protein composition in two lines.

Physiological assays identified no clear differences between 35S::G1543 and wild-type plants.

Potential Applications

The altered levels of chlorophylls, carotenoids, seed oils, and proteins that resulted from overexpression of the gene in *Arabidopsis* indicate that G1543 or its equivalogs or its equivalogs might

used to manipulate the composition of these substances in seed, with applications toward the improvement in the nutritional value of foodstuffs (for example, by increasing lutein).

Enhanced chlorophyll and carotenoid levels could also improve yield in crop plants. For instance lutein, like other xanthophylls such as zeaxanthin and violaxanthin, is an essential component in the protection of the plant against the damaging effects of excessive light. Specifically, lutein contributes, directly or indirectly, to the rapid rise of non-photochemical quenching in plants exposed to high light. Crop plants engineered to contain higher levels of lutein could therefore have improved photo-protection, possibly leading to less oxidative damage and better growth under high light. Additionally, elevated chlorophyll levels might increase photosynthetic capacity.

G1543 or its equivalogs might be applied to modify plant stature. This could be used to produce crops that are more resistant to damage by wind and rain, or more amenable to harvest. Plants with altered stature might also be of interest to the ornamental plant market.

This gene or its equivalogs may also be used to alter oil production in seeds, which may be very important for the nutritional quality and caloric content of foods

G1792 (SEQ ID NO: 331 and 332)

Published Information

G1792 was identified in the sequence of BAC clone K14B15 (AB025608, gene K14B15.14).

Experimental Observations

G1792 (SEQ ID NO: 331) was studied using transgenic plants in which the gene was expressed under the control of the 35S promoter. 35S::G1792 plants were more tolerant to the fungal pathogens *Fusarium oxysporum* and *Botrytis cinerea* and showed fewer symptoms after inoculation with a low dose of each pathogen. This result was confirmed in T2 lines. The effect of G1792 overexpression in increasing tolerance to pathogens received further, incidental confirmation. T2 plants of two 35S::G1792 lines had been growing in a room that suffered a serious powdery mildew infection. For each line, a pot of six plants was present in a flat containing nine other pots of lines from unrelated genes. In either of the two different flats, the only plants that were free from infection were those from the 35S::G1792 line. This observation indicated that G1792 overexpression might be used to increase resistance to powdery mildew. Additional experiments confirmed that 35S::G1792 plants showed increased tolerance to *Erysiphe*. G1792 was ubiquitously expressed, but appeared to be induced by salicylic acid.

35S::G1792 overexpressing plants also showed more tolerance to growth under nitrogen-limiting conditions. In a root growth assay under conditions of limiting N, 35S::G1792 lines were slightly less stunted. In a germination assay that monitored the effect of C on N signaling through anthocyanin production on high sucrose plus and minus glutamine the 35S::G1792 lines made less anthocyanin on high sucrose plus glutamine, suggesting that the gene can be involved in the plants ability to monitor their carbon and nitrogen status.

G1792 overexpressors were more tolerant to drought conditions than wild-type plants in soil-based assays.

G1792 overexpressing plants showed several mild morphological alterations: leaves were dark green and shiny, and plants bolted, subsequently senesced, slightly later than wild-type controls. Among the T1plants, additional morphological variation (not reproduced later in the T2 plants) was observed: many showed reductions in size as well as aberrations in leaf shape, phyllotaxy, and flower development.

Potential Applications

G1792 or its equivalogs can be used to engineer pathogen-resistant plants. In addition, it can also be used to improve seedling germination and performance under conditions of limited nitrogen.

Potential utilities of this gene or its equivalogs also include increasing chlorophyll content allowing more growth and productivity in conditions of low light. With a potentially higher photosynthetic rate, fruits could have higher sugar content. Increased carotenoid content could be used as a nutraceutical to produce foods with greater antioxidant capability.

G1792 or its equivalogs could be used to manipulate wax composition, amount, or distribution, which in turn could modify plant tolerance to drought and/or low humidity or resistance to insects, as well as plant appearance (shiny leaves). In particular, it would be interesting to see what the effect of increased wax deposition on leaves of a plant like cotton would do to drought resistance or water use efficiency. A possible application for this gene might be in reducing the wax coating on sunflower seeds (the wax fouls the oil extraction system during sunflower seed processing for oil). For this purpose, antisense or co-suppression of the gene in a tissue specific manner might be useful

G1816 (SEQ ID NO: 2142 and 2143)

G1816 corresponds to *TRIPTYCHON* (*TRY*), a gene that regulates epidermal cell specification in the leaf and root (Schnittger et al., 1998; 1999; Schellmann et al., 2002). The gene was included in our earlier studies as a putative paralog of G682 and based on the increased resistance of 35S::G1816 lines to osmotic stress conditions such as high levels of glucose.

The aim of this study was to re-assess 35S::G1816 lines and determine whether overexpression of the gene could confer enhanced stress tolerance in a comparable manner to G682. We also sought to examine whether use of a two-component overexpression system would produce any strengthening of the phenotype relative to the use of a 35S direct promoter-fusion.

We have now generated 35S::G1816 lines using the two component system. These lines showed a strong glabrous phenotype, similar to what was observed during our phase I study, and similar to the effect produced by G682 overexpression. However, many of the 35S::G1816 lines were noted to be smaller than controls, an effect that had not been previously recognized.

Ten of the two component lines were tested in plate based physiology assays, and all ten lines showed a strong resistance to osmotic stress when germinated on sucrose plates. All the lines examined

also showed increased density of root hairs. These effects are broadly comparable to those observed in G682 lines, suggesting that the two genes likely have very related functions. However, 35S::G682 lines showed positive results in a greater number of assays than the 35S::G1816 lines (see 35S::G682 report), perhaps indicating that G682 can protect against a greater range of drought related stresses than G1816.

Table 23. G1816 35S, 2-components supTfn

Line	Germ. in high NaCl	Germ. in high mannitol	Sucrose	ABA	Germ in heat	Germ in cold	Growth in heat	Drought	Chilling	G682-like root morph.
304	wt	wt	++	wt	wt	wt	wt	wt	wt	+
306	wt	wt	+	wt	wt	wt	wt	wt	wt	+
311	wt	wt	++	wt	wt	wt	wt	wt	wt	+
313	wt	wt	+	wt	wt	wt	wt	wt	wt	+
325	wt	wt	++	wt	wt	wt	wt	wt	wt	+
327	wt	wt	+	wt	wt	wt	wt	wt	wt	+
345	wt	wt	++	wt	wt	wt	wt	wt	wt	+
350	wt	wt	+	wt	wt	wt	wt	wt	wt	+
351	wt	wt	+	wt	wt	wt	wt	wt	wt	+
353	wt	wt	++	wt	wt	wt	wt	wt	wt	+

Potential Applications

Based on the tolerance of 35S::G1816 lines to osmotic stress, G682 and related sequences such as G1816 are good candidates for use in the alleviation of drought related stress. The strong performance of 35S::G1816 lines on plates containing high levels of sugar particularly suggests that the gene might also be applied to manipulate sugar-sensing responses. The decrease in size seen in some of the lines, suggests that optimization of stress tolerance gene might benefit by use of different promoters or protein modifications.

The epidermal phenotypes seen in 35S::G1816 lines indicate that the gene could also be used to modify developmental characters such as the formation of trichomes or root hairs.

G1820 (SEQ ID NO: 341 and 342)

Published Information

G1820 is a member of the Hap5 subfamily of CCAAT-box-binding transcription factors. G1820 was identified as part of the BAC clone MBA10, accession number AB025619 released by the *Arabidopsis* Genome sequencing project.

Experimental Observations

The complete sequence of G1820 was determined.

In soil-based assays, 35S::G1792 direct fusion overexpressing plants were significantly more drought tolerant than wild-type control plants

The function of this gene was also analyzed using transgenic plants in which G1820 was expressed under the control of the 35S promoter with the two-component transformation system. A wide range of morphological alterations was observed, similar to those seen in our earlier studies.

In plate-based physiology assays on the two-component lines, many of our previously observed drought stress-related phenotypes were confirmed. The majority of lines were tolerant, to varying extents, in salt, mannitol, sucrose, ABA and cold in germination assays.

It should be emphasized that we have observed stress tolerance phenotypes for several other G481 related genes including G482, G485, G1836, and non-*Arabidopsis* sequences. The similar effects seen when these genes are overexpressed strongly suggest that they are likely be functionally related, at least with respect to this phenotype.

Overexpression of G1820 also consistently reduced the time to flowering. Under continuous light conditions at 20-25 C, the 35S::G1820 transformants displayed visible flower buds several days earlier than control plants. The primary shoots of these plants typically started flower initiation 1-4 leaf plastochrons sooner than those of wild type. Such effects were observed in all three T2 populations and in a substantial number of primary transformants.

When biochemical assays were performed, some changes in leaf fames were detected. In one line, an increase in the percentage of 18:3 and a decrease in 16:1 were observed. G1820 overexpressors behaved similarly to wild-type controls in other biochemical assays. As determined by RT-PCR, G1820 was highly expressed in embryos and siliques. No expression of G1820 was detected in the other tissues tested. G1820 expression appeared to be induced in rosette leaves by cold and drought stress treatments, and overexpressing lines showed tolerance to water deficit and high salt conditions.

One possible explanation for the complexity of the G1820 overexpression phenotype is that the gene is involved in the cross talk between ABA and GA signal transduction pathways. It is well known that seed dormancy and germination are regulated by the plant hormones abscisic acid (ABA) and gibberellin (GA). These two hormones act antagonistically with each other. ABA induces seed dormancy in maturing embryos and inhibits germination of seeds. GA breaks seed dormancy and promotes germination. It is conceivable that the flowering time and ABA insensitive phenotypes observed in the G1820 overexpressors are related to an enhanced sensitivity to GA, or an increase in the level of GA, and that the phenotype of the overexpressors is unrelated to ABA. In *Arabidopsis*, GA is thought to be required to promote flowering in non-inductive photoperiods. However, the drought and salt tolerant phenotypes would indicate that ABA signal transduction is also perturbed in these plants. It seems counterintuitive for a plant with salt and drought tolerance to be ABA insensitive since ABA seems to activate signal transduction pathways involved in tolerance to salt and dehydration stresses. One explanation is that ABA levels in the G1820 overexpressors are also high but that the plant is unable to perceive or transduce the signal.

G1820 overexpressors also had decreased seed oil content and increased seed protein content compared to wild-type plants

Table 24. G1820 35S, 2-components-supTfn

Line	Germ. in high NaCl	Germ. in high mannitol	Sucrose	ABA	Germ. in heat	Germ. in cold	Growth in heat	Drought	Chilling
305	wt	wt	+	++	wt	+	wt	wt	+
310	wt	wt	wt	+	wt	wt	wt	wt	wt
321	wt	wt	+	+	wt	+	wt	wt	wt
323	wt	wt	wt	wt	wt	wt	wt	wt	wt
325	wt	wt	wt	+	wt	wt	wt	wt	+
326	wt	wt	+	+	wt	wt	wt	wt	wt
327	wt	wt	wt	wt	wt	wt	wt	wt	wt
341	++	+	+	++	wt	+	wt	wt	wt
343	++	+	+	++	wt	wt	wt	wt	wt
352	++	+	+	+	wt	wt	wt	wt	wt

Potential Applications

G1820 affects ABA sensitivity, and thus when transformed into a plant this transcription factor or its equivalogs may diminish cold, drought, oxidative and other stress sensitivities, and also be used to alter plant architecture, and yield.

The osmotic stress and cold assay results indicate that G481 and related sequences, including G1820, could be used to alter a plant's response to water deficit and can be used to engineer plants with enhanced tolerance to drought, salt stress, and freezing.

G1820 or its equivalogs may also be used to increase a plant's tolerance to cold.

G1820 or its equivalogs could also be used to accelerate flowering time.

G1820 or its equivalogs may be used to modify levels of saturation in oils.

G1820 or its equivalogs may be used to seed protein content.

The promoter of G1820 could be used to drive seed-specific gene expression.

Potential Applications

G1820 or equivalog overexpression may be used to alter seed protein content, which may be very important for the nutritional value and production of various food products

G1836 (SEQ ID NO: 343 and 344)

Published Information

G1836 was identified in the sequence of BAC F14I23, GenBank accession number AC007399, released by the *Arabidopsis* Genome Initiative.

Experimental Observations

The complete sequence of G1836 was determined. The function of this gene was analyzed using transgenic plants in which G1836 was expressed under the control of the 35S promoter. Morphologically,

the plants were somewhat paler than the wild-type controls. This observation did not translate into a detectable difference in the chlorophyll a or chlorophyll b content in these transgenics (see biochemistry data). Overexpression of G1836 affected the plants' ability to tolerate high concentrations of salt in a germination assay. All of the lines showed greater expansion of the cotyledons when seeds are germinated on MS media containing high concentrations of NaCl, indicating they had more tolerance to salt stress compared to the wild-type controls. There was no enhanced tolerance to high salt in older seedlings in a root growth assay. This was not unexpected because salt tolerance in the two developmental stages in often uncoupled in nature indicating mechanistic differences.

G1836 overexpression also resulted in plants that were more drought tolerant than wild-type control plants.

Expression of G1836 was also repressed by *Erysiphe orontii* infection.

Seven of ten lines tested in a recent series of plate-based assays showed enhanced abiotic stress tolerance, as indicated in Table 25. These results included six lines with improved salt tolerance, and several of the lines showed altered sugar sensing in sucrose- and mannitol-based assays, less sensitivity to ABA, and improved tolerance to cold in germination and chilling assays.

Table 25. G1836 35S, 2 components-supTfn

Line	Germ. in high NaCl	Germ. in high mannitol	Sucrose	ABA	Germ. in heat	Germ. in cold	Growth in heat	Drought	Chilling
306	+	wt	+	++	wt	wt	wt	wt	+
362	wt	wt	wt	wt	wt	wt	wt	wt	wt
363	wt	wt	wt	wt	wt	wt	wt	wt	wt
365	wt	wt	wt	wt	wt	wt	wt	wt	wt
367	wt	wt	wt	wt	wt	++	wt	wt	+
381	+	wt	+	+	wt	wt	wt	wt	+
383	+	wt	wt	wt	wt	wt	wt	wt	+
384	+	wt	+	+	wt	wt	wt	wt	wt
385	+	+	+	wt	wt	wt	wt	wt	wt
386	+	wt	+	+	wt	wt	wt	wt	wt

Potential Applications

The results of these abiotic stress assays indicate that G1836 can be used to increase plant tolerance to drought, soil salinity and cold conditions during germination or at the seedling stage. The results of these studies confirm our earlier conclusions that G481 and its related genes, including G1836, are excellent candidates for improvement of drought and cold-related stress tolerance in plants.

G1930 (SEQ ID NO: 369 and 370)

Published Information

G1930 was identified in the sequence of P1 clone K13N2 (gene K13N2.7, GenBank protein accession number BAA95760).

Experimental Observations

G1930 was ubiquitously expressed and did not appear to be induced by any of the conditions tested.

We generated 35S::G1930 lines via both the direct promoter-fusion and the 2-component methods. Both types of lines exhibited a variety of morphological phenotypes including reduced size, slow growth, and alterations in leaf orientation. In some lines, changes in leaf shape, hypocotyl length, trichome density, flowering time and non-specific floral abnormalities that reduced fertility were also observed.

We tested both two component and direct promoter-fusion lines under a variety of plate based treatments. Comparable results were obtained with both types of lines, but the 2-component lines generally showed stronger phenotypes, suggesting perhaps, that this system afforded an amplification of G1930 activity relative to the direct promoter-fusion. All twenty of the lines tested gave positive results in one or more of the stress treatments, including, sucrose, NaCl, ABA, cold germination and cold growth. Particularly strong resistance was seen in the NaCl and sucrose germination assays.

An increase in the amount of chlorophylls a and b in seeds of two T2 lines was detected.

We have obtained comparable developmental effects as well as a strong enhancement of drought related stress tolerance from all four of the *Arabidopsis* genes in the G867 study group (G9, G867, G993, and G1930). The almost identical phenotypic effects produced by these genes strongly suggest that they are functionally equivalent.

Table 26. G1930 35S, Direct promoter-fusion and 2-components-supTfn

Line	Transfor-mation	Germ. in high NaCl	Germ. in high mannitol	Sucrose	ABA	Germ in heat	Germ in cold	Growth in heat	Drought	Chilling
304	Direct fusion	+	wt	wt	wt	wt	wt	wt	wt	wt
305	Direct fusion	wt	wt	++	wt	wt	+	wt	wt	wt
306	Direct fusion	+	wt	wt	wt	wt	wt	wt!	wt	+
308	Direct fusion	wt	wt	++	wt	wt	wt	wt	wt	+
309	Direct fusion	wt	wt	wt	wt	wt	wt	wt	wt	+
311	Direct fusion	wt	wt	+	wt	wt	+	wt	wt	+
365	Direct fusion	+	wt	++	wt	wt	wt	wt	wt	wt
367	Direct fusion	+	wt	++	wt	wt	wt	wt	wt	+
369	Direct fusion	+	wt	wt	wt	wt	wt	wt	wt	+
370	Direct fusion	+	wt	wt	wt	wt	wt	wt	wt	+

321	2-comp	+	wt	++	wt	wt	wt	wt	wt	+
322	2-comp	+	wt	+	+	wt	wt	wt	wt	+
324	2-comp	+	wt	+	+	wt	wt	wt	wt	+
327	2-comp	+	wt	+	wt	wt	wt	wt	wt	wt
329	2-comp	+	wt	+	wt	wt	wt	wt	wt	+
331	2-comp	+	wt	++	+	wt		wt	wt	+
332	2-comp	+	wt	+	+	wt	wt	wt	wt	wt
334	2-comp	+	wt	wt	wt	wt	wt	wt	wt	wt
336	2-comp	+	wt	++	+	wt	wt	wt	wt	+
339	2-comp	+	wt	wt	wt	wt	wt	wt	wt	wt
331	2-comp						+			

Potential Applications

Based on the results of our overexpression studies, G867 and related sequences such as G1930 are excellent candidate genes for improvement of abiotic stress tolerance in commercial plant species. The morphological effects associated with their overexpression indicate that tissue specific or conditional promoters might be used to optimize the utility of these genes.

This gene could also be used to regulate the levels of chlorophyll in seeds.

G2053 (SEQ ID NO: 389 and 390)

Published Information

G2053 was identified in the sequence of BAC T27C4, GenBank accession number AC022287, released by the *Arabidopsis* Genome Initiative.

Experimental Observations

The function of G2053 was analyzed using transgenic plants in which the gene was expressed under the control of the 35S promoter. In a root growth assay on media containing high concentrations of PEG, G2053 overexpressors showed more root growth compared to wild-type controls. G2053 overexpressors also were significantly more drought tolerant than wild-type control plants.

Potential Applications

G2053 or its equivalogs could be used to alter a plant's response water deficit conditions and, therefore, could be used to engineer plants with enhanced tolerance to drought, salt stress, and freezing.

G2133 (SEQ ID NO: 407 and 408)

G2133 corresponds to gene F26A9.11 (AAF23336).

Experimental Observations

The function of G2133 was studied using transgenic plants in which the gene was expressed under the control of the 35S promoter.

G2133 expression was detected in a variety of tissues: flower, leaf, embryo, and silique samples. Its expression might be altered by several conditions, including auxin treatment, osmotic stress, and *Fusarium* infection. Overexpression of G2133 caused a variety of alterations in plant growth and development: delayed flowering, altered inflorescence architecture, and a decrease in overall size and fertility. G2133 plants were more tolerant to glyphosate than wild-type control plants.

At early stages, 35S::G2133 transformants were markedly smaller than controls and displayed curled, dark-green leaves. Most of these plants remained in a vegetative phase of development substantially longer than controls, and produced an increased number of leaves before bolting. In the most severely affected plants, bolting occurred more than a month later than in wild type (24-hour light). In addition, the plants displayed a reduction in apical dominance and formed large numbers of shoots simultaneously, from the axils of rosette leaves. These inflorescence stems had short internodes, and carried increased numbers of cauline leaf nodes, giving them a very leafy appearance. The fertility of 35S::G2133 plants was generally very low. In addition, G2133 overexpressing lines were found to be more resistant to the herbicide glyphosate in initial and repeat experiments.

G2133 is a paralog of G47, the latter having been known from earlier studies to confer a drought tolerance phenotype when overexpressed. It was thus not surprising when G2133 was also shown to induce drought tolerance in a number of 35S::G2133 lines challenged in soil-based drought assays. After re-watering, all of the plants of both G2133 overexpressor lines became reinvigorated, and all of the control plants died or were severely affected by the drought treatment.

Potential Applications

G2133 and its equivalogs can be used to increase the tolerance of plants to drought and to other osmotic stresses. G2133 could also be used for the generation of glyphosate resistant plants, and to increase plant resistance to oxidative stress and glyphosate.

G2153 (SEQ ID NO: 417 and 418)

Published Information

The sequence of G2153 was obtained from *Arabidopsis* genomic sequencing project, GenBank accession number AC011437, based on its sequence similarity within the conserved domain to other AT-hook related proteins in *Arabidopsis*. G2153 corresponds to gene F7O18.4 (AAF04888).

Experimental Observations

The complete sequence of G2153 was determined. G2153 was strongly expressed in roots, embryos, siliques, and germinating seed, but at low or undetectable levels in shoots, flowers, and rosette leaves. It was not significantly induced or repressed by any condition tested.

The function of this gene was analyzed using transgenic plants in which G2153 was expressed under the control of the 35S promoter. Overexpression of G2153 in *Arabidopsis* resulted in seedlings

with an altered response to osmotic stress. In a germination assay on media containing high sucrose, G2153 overexpressors had more expanded cotyledons and longer roots than the wild-type controls. This phenotype was confirmed in repeat experiments on individual lines, and all three lines showed osmotic tolerance. Increased tolerance to high sucrose could also be indicative of effects on sugar sensing. Overexpression of G2153 produced no consistent effects on *Arabidopsis* morphology, and no altered phenotypes were noted in any of the biochemical assays.

Potential Applications

G2153 or its equivalogs can be improve a plant's response to drought, salt stress, and freezing. G2153 or its equivalogs may also be useful for altering a plant's response to sugars.

G2155 (SEQ ID NO: 419 and 420)

Published Information

The sequence of G2155 was obtained from *Arabidopsis* genomic sequencing project, GenBank accession number AC012188.

Experimental Observations

The complete sequence of G2155 was determined. G2155 expression was detected at low levels only in flowers and embryos. It was not induced in rosette leaves by any condition tested.

The function of this gene was analyzed using transgenic plants in which G2155 was expressed under the control of the 35S promoter. Overexpression of G2155 produced a marked delay in the time to flowering. Under continuous light conditions, 35S::G2155 transformants displayed a considerable extension of vegetative development, and typically formed flower buds about two weeks later than wild-type controls. At early stages, the plants were slightly small and had rather rounded leaves compared to wild type. However, later in development, when the leaves were fully expanded, 35S::G2155 plants became very large, dark-green, and senesced much later than controls.

In addition, overexpression of G2155 resulted in an increase in seed glucosinolate M39497 in two T2 lines. No other phenotypic alterations were observed in any of the biochemical or physiological assays.

Potential Applications

G2155 or equivalog overexpression may be used to delay flowering.

G2155 or its equivalogs could also be used to alter seed glucosinolate composition.

G2345 (SEQ ID NO: 2171 and 2172)

G2345 is a putative paralog of G481, and is a member of the HAP3 subgroup of the CCAAT transcription factor family. During our earlier program, we found that G2345-overexpressing plants were

similar to wild-type plants in all assays performed. The aim of this study was to reassess the role of G2345 in drought stress related tolerance via two-component overexpression.

We have now generated 35S lines for G2345 using the two component system; three batches of T1 lines were obtained, including lines 301-302, 341-347, and 381-400. Some size variation was apparent in the second batch of plants (341-347), but otherwise, no consistent differences in morphology were observed compared to wild-type controls.

Two of two lines of overexpressors tested showed better germination in cold conditions than did wild-type control plants.

Table 27. G2345 35S, 2-components supTfn

Line	Germ. in high NaCl	Germ. in high mannitol	Sucrose	ABA	Germ. in heat	Germ. in cold	Growth in heat	Drought	Chilling
301	wt	wt	wt	wt	wt		wt	wt	
302	wt	wt	wt	wt	wt		wt	wt	
341	wt	wt	wt	wt	wt		wt	wt	
386	wt	wt	wt	wt	wt		wt	wt	
387	wt	wt	wt	wt	wt		wt	wt	
389	wt	wt	wt	wt	wt		wt	wt	
390	wt	wt	wt	wt	wt		wt	wt	
392	wt	wt	wt	wt	wt		wt	wt	
393	wt	wt	wt	wt	wt	+	wt	wt	
400	wt	wt	wt	wt	wt	+	wt	wt	

Potential Applications

The results of the cold germination assay confirm that G481 and its related sequences, including G2345, are excellent candidates for improving abiotic stress tolerance in plants.

G2509 (SEQ ID NO: 439 and 440)

Published Information

G2509 corresponds to gene T2I1_20 (CAB87920).

Experimental Observations

G2509 (SEQ ID NO: 439) was studied using transgenic plants in which the gene was expressed under the control of the 35S promoter. Overexpression of G2509 caused multiple alterations in plant growth and development, most notably, altered branching patterns, and a reduction in apical dominance, giving the plants a shorter, bushier stature than wild type. Twenty 35S::G2509 primary transformants were examined; at early stages of rosette development, these plants displayed a wild-type phenotype. However, at the switch to flowering, almost all T1 lines showed a marked loss of apical dominance and large numbers of secondary shoots developed from axils of primary rosette leaves. In the most extreme cases, the shoots had very short internodes, giving the inflorescence a very bushy appearance. Such shoots were often very thin and flowers were relatively small and poorly fertile. At later stages, many

plants appeared very small and had a low seed yield compared to wild type. In addition to the effects on branching, a substantial number of 35S::G2509 primary transformants also flowered early and had buds visible several days prior to wild type. Similar effects on inflorescence development were noted in each of three T2 populations examined. The branching and plant architecture phenotypes observed in 35S::G2509 lines resemble phenotypes observed for three other AP2/EREBP genes: G865, G1411, and G1794, G2509, G865, and G1411 form a small clade within the large AP2/EREBP family, and G1794, although not belonging to the clade, is one of the AP2/EREBP genes closest to it in the phylogenetic tree. It is thus likely that all these genes share a related function, such as affecting hormone balance.

G2509 overexpressing plants had increased seed protein compared to wild-type control plants.

Overexpression of G2509 in *Arabidopsis* resulted in an increase in alpha-tocopherol in seeds in two T2 lines. G2509 was ubiquitously expressed in *Arabidopsis* plant tissue. G2509 expression levels were altered by a variety of environmental or physiological conditions.

Potential Applications

G2509 or its equivalogs can be used to manipulate plant architecture and development, alter tocopherol composition, and flowering time.

G2583 (SEQ ID NO: 449 and 450)

Published Information

G2583 corresponds to gene F2I11_80 (CAB96654).

Experimental Observations

35S::G2583 plants exhibited extremely glossy leaves. At early stages, 35S::G2583 seedlings appeared normal, but by about two weeks after sowing, the plants exhibited very striking shiny leaves, which were apparent until very late in development. Many lines displayed a variety of other effects such as a reduction in overall size, narrow curled leaves, or various non-specific floral abnormalities, which reduced fertility. These effects on leaf appearance were observed in 18 of 20 primary transformants, and in all the plants from 4 of 6 T2 lines. The glossy nature of the leaves may be a consequence of changes in epicuticular wax content or composition. G2583 belongs to a small clade within the large AP2/EREBP *Arabidopsis* family that also contains G975, G1387, and G977. G975 overexpression causes a substantial increase in leaf wax and a morphology resembling that of 35S::G2583 plants. G2583 was ubiquitously expressed at higher levels in root, flower, embryo, and siliques.

Potential Applications

G2583 or its equivalogs can be used to modify plant appearance by producing shiny leaves. In addition, it or its equivalogs can be used to manipulate wax composition, amount, or distribution, which in turn can modify plant tolerance to drought and/or low humidity or resistance to insects.

G2718 (SEQ ID NO: 2191 and 2192)

G2718 was included in our earlier studies as a putative paralog of G682. A genetic analysis of G2718 function has not yet been published. The aim of this study was to re-assess 35S::G2718 lines and determine whether overexpression of the gene could confer enhanced stress tolerance in a comparable manner to G682. We also sought to examine whether use of a two-component overexpression system would produce any strengthening of the phenotype relative to the use of a 35S direct promoter-fusion.

We have now generated 35S::G2718 lines using the two component system. These lines showed a strong glabrous phenotype, similar to what was observed during our earlier studies, and similar to the effect produced by G682 overexpression. Almost all of the lines tested were more tolerant to high sucrose in an osmotic stress assay, and two lines were found to be insensitive to ABA.

Table 28. G1816 35S, 2-components supTfn

Line	Germ. in high NaCl	Germ. in high mannitol	Sucrose	ABA	Germ in heat	Germ in cold	Growth in heat	Drought	Chilling	G682-like root morph.
341	wt	wt	wt	+	wt		wt	wt		wt
342	wt	wt	+	+	wt		wt	wt		+
425	wt	wt	+	wt	wt		wt	wt		+
427	wt	wt	+	wt	wt		wt	wt		+
428	wt	wt	wt	wt	wt		wt	wt		+
429	wt	wt	+	wt	wt		wt	wt		+
431	wt	wt	+	wt	wt		wt	wt		+
432	wt	wt	+	wt	wt		wt	wt		++
433	wt	wt	+	wt	wt		wt	wt		wt
439	wt	wt	+	wt	wt		wt	wt		+

Potential Applications

Based on the tolerance of 35S::G1816 lines to osmotic stress, G682 and related sequences such as G1816 are good candidates for use in the alleviation of drought related stress. The strong performance of 35S::G1816 lines on plates containing high levels of sugar particularly suggests that the gene might also be applied to manipulate sugar-sensing responses. However, the decrease in size seen in some of the lines, suggests that the gene might require optimization by use of different promoters or protein modifications, prior to product development.

The epidermal phenotypes seen in 35S::G1816 lines indicate that the gene could also be used to modify developmental characters such as the formation of trichomes or root hairs.

Rice G3377 (SEQ ID NO: 1221 and 2934)

G3377 is a rice gene that was identified as being a putative ortholog of G912. The aim of this project was to determine whether G3377 has an equivalent function to G912 via the analysis of 35S::G3377 *Arabidopsis* lines.

35S::G3377 overexpressors were obtained at relatively low frequency. The lines that were recovered showed a number of striking morphological effects including a reduction in size, dark coloration and delayed flowering. Such features were somewhat comparable to those shown by 35S::G912 lines, indicating that the two genes likely have a similar function.

Lines of plants overexpressing these rice sequences were shown to have increased germination in high salt, mannitol, sucrose and hot conditions, as seen in the following table.

Table 29. G3377 35S, Direct promoter-fusion

Line	Germ. in high NaCl	Germ. in high mannitol	Sucrose	ABA	Germ. in heat	Germ. in cold	Growth in heat	Drought	Chilling
382	+	+	+	wt	wt				
383	wt	+	+	wt	wt				
384	wt	wt	wt	wt	+				
385	wt	wt	wt	wt	wt				
386	wt	wt	+	wt	wt				
388	wt	wt	wt	wt	wt				
389	wt	wt	wt	wt	+				
390	wt	wt	wt	wt	wt				
391	wt	wt	wt	wt	wt				
392	wt	wt	wt	wt	wt				

Potential Applications

Given the comparable morphological effects of G3377 and G912 overexpression, it is probable that the genes have comparable roles.

The delayed flowering exhibited by 35S::G3377 lines suggest that the gene might also be applied to modify flowering time; in particular, an extension of vegetative growth can significantly increase biomass and result in substantial yield increases. In some species (for example sugar beet), where the vegetative parts of the plant constitute the crop, it would be advantageous to delay or suppress flowering in order to prevent resources being diverted into reproductive development. Additionally, delaying flowering beyond the normal time of harvest could alleviate the risk of transgenic pollen escape from such crops.

The results of the abiotic stress assays confirm that G912 and its related sequences, including the rice G3377 sequence, are excellent candidates for improving abiotic stress tolerance in plants.

Rice G3392 (SEQ ID NO: 1082 and 2920)

G3392 is a rice gene that was identified as being a putative ortholog of G682. The aim of this project was to determine whether G3392 has an equivalent function to the G682-related genes from *Arabidopsis* via the analysis of 35S::G3392 *Arabidopsis* lines.

We have now generated 35S::G3392 lines; these plants showed comparable morphological effects to 35S::G682 lines and exhibited a glabrous phenotype combined with a reduction in overall size.

Such similarities in phenotypes suggest that the genes have similar functions. Interestingly, many of the 35S::G3392 lines also produced pale yellow seed, which likely indicated a reduction in anthocyanin levels in the seed coat. Such an effect was not observed 35S::G682 seed, but G682 and its paralogs were found during our genomics studies to inhibit anthocyanin production.

Table 30. G3392 35S, Direct promoter-fusion

Line	Germ in high NaCl	Germ. in high mannitol	Sucrose	ABA	Germ in heat	Germ in cold	Heat growth	Drought	Chilling	G682-like root morph.
301	wt	wt	wt	wt	wt	+	wt	+	+	+
305	wt	wt	+	wt	wt	+	wt	+	+	+
306	+	wt	+	wt	wt	+	-	+	+	+
321	wt	wt	+	wt	wt	+	wt	+	+	+
322	+	wt	wt	wt	wt	+	-	ND	+	+
341	wt	+	+	wt	wt	+	wt	+	+	+
342	+	+	++	wt	wt	+	wt	+	+	+
346	wt	+	+	wt	wt	+	wt	+	+	+
347	wt	wt	wt	wt	wt	+	wt	+	+	+
348	wt	wt	+	wt	wt	+	wt	+	+	+

Potential Applications

The results of these abiotic stress assays confirm that G682 and its related sequences, including the rice G3392 sequence, are excellent candidates for improving abiotic stress tolerance in plants.

The effect of G3392 on epidermal patterning indicates that the gene could be applied to manipulate trichome development; in some species trichomes accumulate valuable secondary metabolites and in other instances are thought to provide protection against predation. The lighter coloration of 35S::G3392 plants could indicate that G3392 might be used to regulate the production of flavonoid related compounds, which contribute to the nutritional value of foodstuffs.

Rice G3393 (SEQ ID NO: 559 and 2921)

G3393 is a putative rice ortholog of G682. 35S::G3393 lines plants showed comparable morphological effects to 35S::G682 lines and exhibited a glabrous phenotype combined with a reduction in overall size. These similarities in phenotypes suggest that the genes have similar functions. Many of the 35S::G3393 lines also produced pale yellow seed, which likely indicated a reduction in anthocyanin levels in the seed coat. Such an effect was not observed 35S::G682 seed, but we did find that G682 and its paralogs inhibited anthocyanin production.

Table 31. G3393 35S, Direct promoter-fusion

Line	Germ in high NaCl	Germ. in high mannitol	Sucrose	ABA	Germ in heat	Germ in cold	Heat growth	Drought	Chilling	G682-like root morph.
305	wt	wt	+	wt	wt	wt		wt	++	+
307	wt	wt	wt	wt	wt	wt		wt	+	+

308	wt	wt	wt	wt	wt	wt		wt	++	+
323	wt	wt	wt	wt	wt	wt	wt	wt	+	+
324	wt	wt	wt	wt	wt	wt	wt	wt	+	+
326	wt	wt	wt	wt	wt	wt	wt	wt	+	+
327	wt	wt	wt	wt	wt	wt	wt	wt	++	+
328	wt	wt	wt	wt	wt	wt	wt	wt	++	+
331	wt	wt	wt	wt	wt	wt	wt	wt	wt	+
333	wt	wt	wt	wt	wt	wt	wt	wt	+	+

Potential Applications

The results of the cold germination and sucrose assays confirm that G682 and its related sequences, including the rice G3393 sequence, are excellent candidates for improving abiotic stress tolerance in plants.

The effect of G3393 on epidermal patterning indicates that the gene could be applied to manipulate trichome development; in some species trichomes accumulate valuable secondary metabolites and in other instances are thought to provide protection against predation. The lighter coloration of 35S:: G3393 plants could indicate that G3393 might be used to regulate the production of flavonoid related compounds, which contribute to the nutritional value of foodstuffs.

Rice G3395 (SEQ ID NO: 790 and 2910)

G3395 is an ortholog of G481 from *Oryza sativa*. G3395 is a member of the HAP3 subgroup of the CCAAT-box binding transcription factor family and corresponds to OsHAP3A and has been recently been shown to influence chloroplast biogenesis (Miyoshi et al. (2003) *Plant J.* 36: 532-540). We have previously indicated that overexpression of G3395 conferred a salt tolerant phenotype in *Arabidopsis* plants (U.S. Patent Application No. 10/675,852, filed September 30, 2003).

Experimental Observations

The aim of the present study was to assess the role of G3395 in drought stress-related tolerance via overexpression, and compare the effects with that of the other G481 orthologs and paralogs.

The majority of 35S::G3395 plants showed a very slight acceleration in flowering time, but a single line showed slightly late flowering. One G3395-overexpressing line was shown to have improved tolerance to drought in a plate-based assay. The same line also showed increased tolerance to high salt concentration relative to wild-type or non-transformed controls, confirming the prior observed result.

Table 32. G3395 35S, Direct promoter-fusion

Line	Germ. in high NaCl	Germ. in high mannitol	Sucrose	ABA	Germ. in heat	Germ in cold	Growth in heat	Drought	Chilling
301	wt	wt	wt	wt	wt	wt	wt	wt	wt
302	+	wt	wt	wt	wt	wt	wt	+	wt
303	wt	wt	wt	wt	wt	wt	wt	wt	wt

304	wt	wt	wt	wt	wt	wt	wt	wt	wt
305	wt	wt	wt	wt	wt	wt	wt	wt	wt
306	wt	wt	wt	wt	wt	wt	wt	wt	wt
307	wt	wt	wt	wt	wt	wt	wt	wt	wt
308	wt	wt	wt	wt	wt	wt	wt	wt	wt
309	wt	wt	wt	wt	wt	wt	wt	wt	wt
310	wt	wt	wt	wt	wt	wt	wt	wt	wt

Potential Applications

The results of this drought assay confirm that G481 and its related sequences, including the rice sequence G3395, are excellent candidates for improving abiotic stress tolerance in plants.

Rice G3397 (SEQ ID NO: 794 and 2911)

G3397, an ortholog of G481 from *Oryza sativa*, is a member of the HAP3 subgroup of the CCAAT-box binding transcription factor family. This gene is phylogenetically most closely related to G485, another member of the HAP3 subgroup. G3397 corresponds to OsHAP3C and has been recently been shown to influence chloroplast biogenesis (Miyoshi et al. (2003) *supra*).

Experimental Observations

The aim of this study was to assess the role of G3397 in drought stress-related tolerance via overexpression, and compare the effects with that of the other G481 orthologs and paralogs.

35S::G3397 plants showed a 1-2 week acceleration in flowering time, compared to wild-type or non-transformed plants. This same phenotype was also noted for the most closely related *Arabidopsis* gene G485. These early flowering lines also were smaller than controls. Ten lines were tested in plate-based physiology assays. One of these lines showed insensitivity to ABA, and another germinated better than wild-type or non-transformed plants in hot conditions.

Table 33. G3397 35S, Direct promoter-fusion

Line	Germ. in high NaCl	Germ. in high mannitol	Sucrose	ABA	Germ. in heat	Germ. in cold	Growth in heat	Drought	Chilling
322	wt	wt	wt	wt	wt				
323	wt	wt	wt	wt	wt				
324	wt	wt	wt	wt	wt				
326	wt	wt	wt	+	wt				
328	wt	wt	wt	wt	wt				
330	wt	wt	wt	wt	wt				
334	wt	wt	wt	wt	wt				
335	wt	wt	wt	wt	wt				
338	wt	wt	wt	wt	wt				
339	wt	wt	wt	wt	+				

Potential Applications

The results of the heat and ABA germination assays confirm that G481 and its related sequences, including the rice sequence G3397, are excellent candidates for improving abiotic stress tolerance in plants.

Rice G3398 (SEQ ID NO: 796 and 2912)

G3398 from *Oryza sativa* is related to G481, and phylogenetically most closely related to G485. Like G481 and G485, G3398 is a member of the HAP3 subgroup of the CCAAT-box binding transcription factor family.

Experimental Observations

The aim of this study was to assess the role of G3398 in drought stress-related tolerance via overexpression, and compare the effects with that of the other G481 orthologs and paralogs.

35S::G3398 plants showed a 1-2 week acceleration in flowering time, compared to wild-type or non-transformed plants. This same phenotype was also noted for the most closely related *Arabidopsis* gene, G485. These early flowering lines also were smaller than controls.

In the most recent study, six lines overexpressing G3398 have thus far been tested in physiological assays. One line was shown to germinate better in heat than wild-type or non-transformed controls.

Table 34. G3398 35S, Direct promoter-fusion

Line	Germ. in high NaCl	Germ. in high mannitol	Sucrose	ABA	Germ. in heat	Germ. in cold	Growth in heat	Drought	Chilling
303	wt	wt	wt	wt	+				
323	wt	wt	wt	wt	wt				
324	wt	wt	wt	wt	wt				
329	wt	wt	wt	wt	wt				
332	wt	wt	wt	wt	wt				
335	wt	wt	wt	wt	wt				

Potential Applications

The results of the heat germination assay confirm that G481 and its related sequences, including the rice G3398 sequence, are excellent candidates for improving abiotic stress tolerance in plants.

Rice G3399 (SEQ ID NO: 1399 and 2937)

G3399 identified as a putative rice ortholog of G1073. Phylogenetic analysis identifies G3399 along with G3400 as being the most closely related orthologs to G1073. The aim of this project is to determine whether overexpression of G3399 in *Arabidopsis* produces comparable effects to those of G1073 overexpression.

35S::G3399 lines have been obtained containing either of two different constructs. Both constructs produced similar morphological phenotypes; many of the lines were small at early stages, showed alterations in leaf shape, and had slightly delayed flowering. However a significant number of lines developed enlarged lateral organs (leaves and flowers), particularly at later stages.

It is noteworthy that one of the constructs (P21269) contained an amino acid conversion (proline to a glutamine at residue 198, in a conserved domain) relative to the native protein. Lines for this mutated protein showed fewer undesirable morphologies than the wild type version.

The morphologically similar effects caused by overexpression of this rice gene versus G1073 and the *Arabidopsis* paralogs, suggest that they likely have related functions.

Table 35. G3399 35S, Direct promoter-fusion

Line	Germ. in high NaCl	Germ. in high mannitol	Sucrose	ABA	Germ. in heat	Germ. in cold	Growth in heat	Drought	Chilling
321	wt	wt	wt	wt	wt	wt	wt	wt	wt
322	wt	wt	wt	wt	wt	wt	wt	wt	wt
323	wt	wt	wt	wt	wt	wt	wt	wt	wt
325	wt	wt	wt	wt	++	wt	wt	wt	wt
330	wt	wt	wt	wt	wt	wt	wt	wt	wt
331	wt	wt	wt	wt	wt	wt	wt	wt	wt
332	wt	wt	wt	wt	wt	wt	wt	wt	wt
336	wt	wt	wt	wt	wt	wt	wt	wt	wt
338	wt	wt	wt	wt	+	wt	wt	wt	wt
340	wt	wt	wt	wt	wt	wt	wt	wt	wt
347	wt	wt	wt	wt	wt	wt	wt	wt	+
348	wt	wt	wt	wt	wt	wt	wt	wt	+

Potential Applications

The morphological phenotype induced by G3399 indicates that the gene could be used to modify traits such as organ size and flowering time. This study also identified a specific region of the G3399 protein that might be modified in order to optimize the acquisition of desirable phenotypes. In cases where the increased size conferred by G3399 overexpression would be undesirable, the morphological changes caused by G3440 overexpression may optimized by the use of, for example, inducible or tissue specific promoters.

The results of the heat germination and chilling assays confirm that G1073 and its related sequences, including the rice G3399 sequence, are excellent candidates for improving abiotic stress tolerance in plants.

Rice G3429 (SEQ ID NO: 2945 and 2946)

G3429 from *Oryza sativa* is a gene related to G481. From our phylogenetic analysis, G3429 appears to be more distantly related to G481 than the other non-*Arabidopsis* genes in this study (Figure

3). The gene encodes a protein corresponding to OsNF-YB1 and has been shown to form a ternary complex with a MADS protein OsMADS18 (Masiero et al. (2002) *J. Biol. Chem.* 277: 26429-26435).

Experimental Observations

The aim of this project was to assess the role of G3429 in drought stress-related tolerance, and compare the effects with those of the G481 related genes.

Out of twenty 35S::G3429 T1 plants examined, six were notably late flowering and had narrow leaves compared to wild-type or non-transformed plants.

Six of ten G3429-overexpressing lines were more tolerant to high salt conditions than wild-type or non-transformed plants.

Table 36. G3429 35S, Direct promoter-fusion

Line	Germ. in high NaCl	Germ. in high mannitol	Sucrose	ABA	Germ. in heat	Germ. in cold	Growth in heat	Drought	Chilling
301	wt	wt	wt	wt	wt	wt	wt	wt	wt
302	+	wt	wt	wt	-	wt	wt	wt	wt
304	+	wt	wt	wt	wt	wt	wt	wt	wt
305	+	wt	wt	wt	wt	wt	wt	wt	wt
308	wt	wt	wt	wt	wt	wt	wt	wt	wt
310	wt	wt	wt	wt	wt	wt	wt	wt	wt
311	wt	wt	wt	wt	wt	wt	wt	wt	wt
312	+	wt	wt	wt	wt	wt	wt	wt	wt
313	+	wt	wt	wt	wt	wt	wt	wt	wt
319	+	wt	wt	wt	wt	wt	wt	wt	wt

Potential Applications

The results of the heat germination and chilling assays confirm that G481 and its related sequences, including the rice G3429 sequence, are excellent candidates for improving abiotic stress tolerance in plants.

Corn G3431 (SEQ ID NO: 1089, 556 and 2922)

G3431 is a maize gene that was identified as being a putative ortholog of G682. The aim of this project was to determine whether G3431 has an equivalent function to the G682-related genes from *Arabidopsis* via the analysis of 35S::G3431 *Arabidopsis* lines.

We have now generated 35S::G3431 lines; these plants showed comparable morphological effects to 35S::G682 lines and exhibited a glabrous phenotype combined with a reduction in overall size. These similarities in phenotypes suggest that the genes have similar functions. Interestingly, some of the 35S::G3431 lines also produced pale yellow seed, which likely indicated a reduction in anthocyanin levels in the seed coat. Such an effect was not observed 35S::G682 seed, but G682 and its paralogs were found during earlier genomics studies to inhibit anthocyanin production.

Table 37. G3431 35S, Direct promoter-fusion

Line	Germ in high NaCl	Germ. in high mannitol	Sucrose	ABA	Germ in heat	Germ in cold	Heat growth	Drought	Chilling	G682-like root morph.
303	wt	wt	wt	wt	+	wt	wt	wt	++	+
306	wt	wt	wt	wt	+	wt	wt	wt	wt	-
321	wt	wt	+	wt	wt	wt	wt	wt	+	+
322	wt	wt	wt	wt	wt	wt	wt	wt	+	+
325	wt	wt	+	wt	wt	wt	wt	wt	+	+
327	wt	wt	+	wt	wt	wt	wt	wt	wt	+
328	wt	wt	+	wt	wt	wt	wt	wt	+	+
331	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt
333	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt
340	wt	wt	wt	wt	wt	wt	wt	wt	wt	+

Potential Applications

The results of these abiotic stress assays confirm that G682 and its related sequences, including the corn G3431 sequence, are excellent candidates for improving abiotic stress tolerance in plants.

The effect of G3431 on epidermal patterning indicates that the gene could be applied to manipulate trichome development; in some species trichomes accumulate valuable secondary metabolites and in other instances are thought to provide protection against predation. The lighter coloration of 35S:: G3431 plants could indicate that G3431 might be used to regulate the production of flavonoid related compounds, which contribute to the nutritional value of foodstuffs.

Corn G3434 (SEQ ID NO: 806, and 2913)

G3434 is an ortholog of G481 and G482 from *Zea mays*, and is a member of the HAP3 subgroup of the CCAAT-box binding transcription factor family. Among the *Arabidopsis* paralogs in the G481/G482 study group, G3434 is phylogenetically most closely related to G481.

Experimental Observations

The aim of this study is to assess the role of G3434 in drought stress-related tolerance via overexpression, and compare the effect with that of the other G481/ G482 orthologs and paralogs. As seen in Table 38, 35S::G3434 lines showed enhanced salt tolerance and improved sugar sensing as compared to wild-type or non-transformed plants.

Table 38. G3434 35S, Direct promoter-fusion

Line	Germ. in high NaCl	Germ. in high mannitol	Sucrose	ABA	Germ. in heat	Germ. in cold	Growth in heat	Drought	Chilling
421	wt	wt	wt	wt	wt				
422	+	wt	wt	wt	wt				
423	wt	wt	wt	wt	wt				
424	wt	wt	wt	wt	wt				

426	wt	wt	wt	wt	wt				
429	+	+	+	wt	wt				
432	wt	wt	wt	wt	wt				
434	+	+	+	wt	wt				
435	wt	wt	wt	wt	wt				
436	wt	wt	wt	wt	wt				

Potential Applications

The results of these salt, mannitol and sucrose stress assays confirm that G481 and its related sequences, including the corn G3434 sequence, are excellent candidates for improving abiotic stress tolerance in plants.

Corn G3436 (SEQ ID NO: 805 and 2914)

G3436 from *Zea mays* is a putative ortholog of G481, and is a member of the HAP3 subgroup of the CCAAT-box binding transcription factor family. Among the *Arabidopsis* paralogs in the G481 study group, this gene is phylogenetically most closely related to G485.

Experimental Observations

The aim of this study was to assess the role of G3436 in drought stress-related tolerance via overexpression, and compare the effects with that of the other G481 orthologs and paralogs.

Twenty lines of 35S::G3436 plants showed accelerated flowering time by about 1 week compared to wild-type or non-transformed plants. This same phenotype was also noted for the most closely related *Arabidopsis* gene, G485. Many of these early flowering lines also were smaller than controls.

As seen in Table 39, G3436 overexpressors performed better than controls in a significant number of assay conditions, including high salt, heat germination, cold germination, and growth in cold.

Table 39. G3436 35S, Direct promoter-fusion

Line	Germ. in high NaCl	Germ. in high mannitol	Sucrose	ABA	Germ. in heat	Germ. in cold	Growth in heat	Drought	Chilling
301	wt	wt	wt	wt	+	+	wt	wt	+
302	wt	wt	wt	wt	wt	wt	wt	wt	wt
304	wt	wt	wt	wt	+	wt	wt	wt	+
305	wt	wt	wt	wt	+	wt	wt	wt	wt
308	wt	wt	wt	wt	wt	wt	wt	wt	wt
309	wt	wt	wt	wt	+	+	wt	wt	wt
312	wt	wt	wt	wt	+	wt	wt	wt	wt
313	wt	wt	wt	wt	wt	wt	wt	wt	wt
314	+	wt	wt	wt	wt	wt	wt	wt	wt
315	wt	wt	wt	wt	+	wt	wt	wt	wt

Potential Applications

The results of the salt, heat, cold germination and chilling assays confirm that G481 and its related sequences, including the corn G3436 sequence, are excellent candidates for improving abiotic stress tolerance in plants.

Corn G3440 (SEQ ID NO: 1246, 1213 and 2936)

G3440 is a maize gene that was identified as being a putative ortholog of G912. The aim of this project was to determine whether G3440 has an equivalent function to G912 via the analysis of 35S::G3440 *Arabidopsis* lines.

35S::G3440 lines were small, dark in coloration, and flowered later than control lines. These effects were very similar to those produced by overexpression of G912 in *Arabidopsis*.

Table 40. G3440 35S, Direct promoter-fusion

Line	Germ. in high NaCl	Germ. in high mannitol	Sucrose	ABA	Germ. in heat	Germ. in cold	Growth in heat	Drought	Chilling
302	wt	wt	wt	wt	wt	wt	wt	wt	wt
306	wt	wt	wt	wt	wt	wt	wt	wt	wt
308	wt	wt	wt	wt	wt	wt	wt	wt	+
309	wt	wt	wt	wt	wt	wt	wt	wt	wt
310	wt	wt	wt	wt	wt	wt	wt	wt	wt
311	wt	wt	wt	wt	wt	wt	wt	wt	wt
312	wt	wt	wt	wt	wt	wt	wt	wt	wt
314	wt	wt	wt	wt	wt	wt	wt	wt	wt
317	wt	wt	wt	wt	wt	wt	wt	wt	wt
320	wt	wt	wt	wt	wt	wt	wt	wt	wt

Potential Applications

Given the comparable morphological effects of G3440 and G912 overexpression, it is probable that the genes have comparable roles. The developmental changes caused by G3440 overexpression suggest that the gene would benefit from optimization with, for example, the use of inducible or tissue specific promoters.

The results of the chilling assay confirm that G912 and its related sequences, including the corn G3440 sequence, are excellent candidates for improving abiotic stress tolerance in plants.

Soy G3445 (SEQ ID NO: 1083 and 2915)

G3445 is a soy gene that was identified as a putative ortholog of G682.

We have now generated 35S::G3445 lines; these plants showed comparable morphological effects to 35S::G682 lines and exhibited a glabrous phenotype combined with a slight reduction in overall size. These similarities in phenotypes suggest that the genes have similar functions.

Table 41. G3445 35S, Direct promoter-fusion

Line	Germ in high NaCl	Germ. in high mannitol	Sucrose	ABA	Germ in heat	Germ in cold	Heat growth	Drought	Chilling	G682-like root morph.
301	wt	wt	wt	+	wt	wt	wt	wt	wt	wt
302	wt	wt	wt	+	wt	wt	wt	wt	wt	+
303	wt	wt	wt	+	wt	wt	wt	wt	wt	wt
321	wt	wt	wt	wt	wt	wt	wt	wt	wt	-
323	wt	wt	wt	+	wt	wt	wt	wt	wt	wt
341	wt	wt	wt	wt	wt	wt	wt	wt	wt	+
342	wt	wt	wt	wt	wt	wt	wt	+	wt	wt
344	wt	wt	wt	wt	wt	wt	wt	wt	wt	+
345	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt
347	wt	wt	wt	wt	wt	wt	wt	+	wt	wt

Potential Applications

The results of the ABA and drought assays, confirm that G682 and its related sequences, including the soy G3445 sequence, are excellent candidates for improving abiotic stress tolerance in plants.

The effect of G3445 on epidermal patterning indicates that the gene could be applied to manipulate trichome development; in some species trichomes accumulate valuable secondary metabolites and in other instances are thought to provide protection against predation.

Soy G3448 (SEQ ID NO: 1087, 553 and 2917)

G3448 is a soy gene that was identified as being a putative ortholog of G682. The aim of this project was to determine whether G3448 has an equivalent function to the G682-related genes from *Arabidopsis* via the analysis of 35S::G3448 *Arabidopsis* lines.

We have now generated 35S::G3448 lines; these plants showed comparable morphological effects to 35S::G682 lines and exhibited a glabrous phenotype combined with a reduction in overall size. These similarities in phenotypes suggest that the genes have similar functions. Additionally the 35S::G3448 lines showed a somewhat lighter coloration than controls, perhaps indicating that levels of pigments such as anthocyanins were reduced in leaf tissue.

Table 42. G3448 35S, Direct promoter-fusion

Line	Germ in high NaCl	Germ. in high mannitol	Sucrose	ABA	Germ in heat	Germ in cold	Heat growth	Drought	Chilling	G682-like root morph.
302	wt	wt	wt	wt	wt	wt	wt	wt	+	+
303	wt	wt	wt	wt	wt	wt	wt	wt	wt	+
305	wt	wt	wt	wt	wt	wt	wt	wt	+	+
308	wt	wt	wt	wt	wt	wt	wt	+	wt	+
309	wt	wt	wt	wt	wt	wt	wt	wt	wt	+
310	wt	wt	wt	wt	wt	wt	wt	wt	wt	+
313	wt	wt	wt	wt	wt	wt	wt	wt	wt	+

314	wt	wt	wt	wt	wt	wt	wt	wt	wt	+
315	wt	wt	wt	wt	wt	wt	wt	wt	+	+
317	wt	wt	wt	wt	wt	wt	wt	wt	wt	+

Potential Applications

The results of these drought and chilling stress assays confirm that G682 and its related sequences, including the soy G3448 sequence, are excellent candidates for improving abiotic stress tolerance in plants.

The effect of G3448 on epidermal patterning indicates that the gene could be applied to manipulate trichome development; in some species trichomes accumulate valuable secondary metabolites and in other instances are thought to provide protection against predation. The lighter coloration of 35S::G3448 plants could indicate that G3448 might be used to regulate the production of flavonoid related compounds, which contribute to the nutritional value of foodstuffs.

Soy G3449 (SEQ ID NO: 1088, 554 and 2918)

G3449 is a soy gene that was identified as being a putative ortholog of G682. The aim of this project was to determine whether G3449 has an equivalent function to the G682-related genes from *Arabidopsis* via the analysis of 35S::G3449 *Arabidopsis* lines.

We have now generated 35S::G3449 lines; these plants showed comparable morphological effects to 35S::G682 lines and exhibited a glabrous phenotype combined with a slight reduction in overall size. These similarities in phenotypes suggest that the genes have similar functions. Additionally, 35S::G3449 transformants were distinctly paler than wild-type at the seedling stage, perhaps indicating a reduction in the levels of pigments such as anthocyanins.

Table 43. G3449 35S, Direct promoter-fusion

Line	Germ in high NaCl	Germ. in high mannitol	Sucrose	ABA	Germ in heat	Germ in cold	Heat growth	Drought	Chilling	G682-like root morph.
303	wt	wt	wt	wt	+	+	wt	wt	wt	+
304	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt
305	wt	wt	wt	wt	wt	+	wt	wt	wt	+
306	wt	wt	wt	wt	wt	wt	wt	wt	wt	+
307	wt	wt	wt	wt	wt	wt	wt	wt	wt	+
309	wt	wt	wt	wt	wt	wt	wt	wt	wt	+
310	wt	wt	wt	wt	wt	wt	wt	wt	wt	+
311	wt	wt	wt	wt	wt	+	wt	wt	+	+
312	wt	wt	wt	wt	wt	wt	wt	wt	wt	+
313	+	wt	wt	wt	wt	wt	wt	wt	wt	+

Potential Applications

The results of the salt and cold germination assay confirm that G682 and its related sequences, including the soy G3449 sequence, are excellent candidates for improving abiotic stress tolerance in plants.

The effect of G3449 on epidermal patterning indicates that the gene could be applied to manipulate trichome development; in some species trichomes accumulate valuable secondary metabolites and in other instances are thought to provide protection against predation. The lighter coloration of 35S::G3449 plants could indicate that G3449 might be used to regulate the production of flavonoid related compounds, which contribute to the nutritional value of foodstuffs.

Soy G3450 (SEQ ID NO: 1084, 550, 1076 and 2919)

G3450 is a soy gene that was identified as being a putative ortholog of G682. Based on a phylogenetic tree built using conserved MYB domains, the G3450 protein appears to be more closely related to the G682-clade of *Arabidopsis* genes than any of the other putative orthologs included the study. The aim of this project was to determine whether G3450 has an equivalent function to the G682-related genes from *Arabidopsis* via the analysis of 35S::G3450 *Arabidopsis* lines.

We have now generated 35S::G3450 lines; these plants showed comparable morphological effects to 35S::G682 lines and exhibited a glabrous phenotype combined with a slight reduction in overall size. These similarities in phenotypes suggest that the genes have similar functions. Interestingly, 35S::G3450 lines were slightly pale and some of the lines produced pale yellow seed, which likely indicated a reduction in anthocyanin levels in the seed coat. Such an effect was not observed in 35S::G682 seed, but G682 and its paralogs were found during our genomics studies to inhibit anthocyanin production.

35S::G3450 lines have recently been tested in drought related assays; all of these lines exhibited an increase in root hair density and seven of the ten lines showed an enhanced performance in one or more of the plate-based drought related stress assays.

The comparable morphological and physiological effects obtained in 35S::G3450 lines versus overexpression lines for the G682-related *Arabidopsis* genes, indicates that the G3450 protein has a very similar or equivalent activity to the *Arabidopsis* proteins

Table 44. G3450 35S, Direct promoter-fusion

Line	Germ in high NaCl	Germ. in high mannitol	Sucrosc	ABA	Germ in heat	Germ in cold	Heat growth	Drought	Chilling	G682-like root morph.
301	wt	wt	wt	wt	wt	wt	wt	wt	wt	+
302	wt	wt	wt	wt	+	+	wt	wt	wt	+
303	wt	wt	wt	wt	wt	+	wt	wt	+	+
304	wt	wt	wt	wt	wt	+	+	wt	wt	+
305	wt	wt	wt	wt	wt	wt	wt	wt	wt	+

306	wt	wt	wt	wt	wt	wt	wt	+	wt	+
307	wt	wt	wt	wt	wt	+	+	wt	+	+
313	wt	wt	wt	wt	wt	wt	wt	wt	+	+
315	+	wt	wt	wt	wt	+	+	+	+	+
317	+	wt	wt	wt	wt	+	wt	wt	+	+

Potential Applications

The results of the cold, heat cold and salt germination and growth assays confirm that G682 and its related sequences, including the soy G3450 sequence, are excellent candidates for improving abiotic stress tolerance in plants.

The effect of G3450 on epidermal patterning indicates that the gene could be applied to manipulate trichome development; in some species trichomes accumulate valuable secondary metabolites and in other instances are thought to provide protection against predation. The lighter coloration of 35S:: G3450 plants could indicate that G3450 might be used to regulate the production of flavonoid related compounds, which contribute to the nutritional value of foodstuffs.

Soy G3452 (SEQ ID NO: 1183, 1162 and 2924)

G3452 is a soy gene that was identified as being a putative ortholog of G867. The aim of this project was to determine whether G3452 has an equivalent function to G867 via the analysis of 35S::G3452 *Arabidopsis* lines.

35S::G3452 lines displayed a number of morphological similarities, such as reduced size and alterations in coloration, to those seen in earlier studies with 35S::G867 lines. A number of the 35S::G3452 lines were also slightly early flowering.

Table 45. G3452 35S, Direct promoter-fusion

Line	Germ. in high NaCl	Germ. in high mannitol	Sucrose	ABA	Germ. in heat	Germ. in cold	Growth in heat	Drought	Chilling
304	+	+	+	wt	wt	+	+	+	
305	+	wt	+	wt	wt	+	wt	wt	
310	wt	wt	wt	wt	wt		wt	wt	+
314	+	wt	+	wt	wt		wt	wt	+
316	+	wt	+	wt	wt		wt	wt	+
318	wt	wt	+	wt	wt		wt	wt	+

Potential Applications

The results of these abiotic stress assays confirm that G867 and its related sequences, including the soy G3452 sequence, are excellent candidates for improving abiotic stress tolerance in plants.

The accelerated flowering seen in 35S::G3452 plants indicate that the gene could be used to manipulate flowering time. In particular, shortening generation times would also help speed-up breeding programs, particularly in species such as trees, which typically grow for many years before flowering.

Conversely, it might be possible to modify the activity of G3452 (or its orthologs) to delay flowering in order to aehieve an increase in biomass and yield.

Soy G3465 (SEQ ID NO: 1206 and 1242)

G3465 is a soy gene that was identified as being a putative ortholog of G912. On a phylogenetic tree, this gene appears to be more closely related to G912 and *CBF1-3*, than to the two related *Arabidopsis* genes, G2107 and G2513. The aim of this project was to determine whether G3465 has an equivalent function to G912 via the analysis of 35S::G3465 *Arabidopsis* lines.

Overexpression of G3465 produced deleterious effects in *Arabidopsis*; 35S::G3465 lines were small, dark in coloration and slow growing. Such features were comparable to, and possibly even more severe than those shown by 35S::G912 lines, indicating that the two genes likely have a similar function. One overexpressing line was shown to have increased germination in high mannitol relative to wild-type control plants, and another line was insensitive to ABA.

Table 46. G3465 35S, Direct promoter-fusion

Line	Germ. in high NaCl	Germ. in high mannitol	Sucrose	ABA	Germ. in heat	Germ. in cold	Growth in heat	Drought	Chilling
321	wt	wt	wt	wt	wt				
322	wt	wt	wt	wt	wt				
323	wt	wt	wt	wt	wt				
324	wt	wt	wt	wt	wt				
341	wt	wt	wt	wt	wt				
343	wt	wt	wt	wt	wt				
344	wt	wt	wt	wt	wt				
346	wt	+	wt	wt	wt				
347	wt	wt	wt	wt	wt				
348	wt	wt	wt	+	wt				

Potential Applications

Given the similar morphological effects of G3465 and G912 overexpression, it is probable that the genes have comparable roles. The morphological effects that were apparent in these lines suggest that the utilization of G3465 might benefit from optimization by use of different promoters or protein modifications.

The results of the mannitol and ABA germination assays confirm that G912 and its related sequences, including the soy G3465 sequence, are excellent candidates for improving abiotic stress tolerance in plants.

Soy G3469 (SEQ ID NO: 1210 and 1237)

G3469 is a soy gene that was identified as being a putative ortholog of G912. Morphologically, the overexpressors of G3469 ranged from being somewhat small in size to plants with no consistent differences to control.

Results of plate based assays show that G3469 is, like its ortholog G912, is able to confer abiotic stress tolerance in plants, as indicated by the greater tolerance than wild type control plants of G3469 overexpressors to chilling conditions and germination in high salt.

Table 47. G3469 35S, Direct promoter-fusion

Line	Germ. in high NaCl	Germ. in high mannitol	Sucrose	ABA	Germ. in heat	Germ. in cold	Growth in heat	Drought	Chilling
302	wt	wt	wt	wt	wt	wt	wt	wt	wt
303	+	wt	wt	wt	wt	wt	wt	wt	wt
305	+	wt	wt	wt	wt	wt	wt	wt	wt
309	wt	wt	wt	wt	wt	wt	wt	wt	wt
310	+	wt	wt	wt	wt	wt	wt	wt	wt
311	+	wt	wt	wt	wt	wt	wt	wt	wt
312	+	wt	wt	wt	wt	wt	wt	wt	wt
314	wt	wt	wt	wt	wt	wt	wt	wt	wt
315	wt	wt	wt	wt	wt	wt	wt	wt	+
319	+	wt	wt	wt	wt	wt		wt	+
304	wt	wt	wt	wt	wt	wt	wt	wt	+
307	wt	wt	wt	wt	wt	wt	wt	wt	+
317	wt	wt	wt	wt	wt	wt	wt	wt	+
318	wt	wt	wt	wt	wt	wt	wt	+	wt

Potential Applications

The results of the salt, drought and chilling assays confirm that G912 and its related sequences, including the soy G3469 sequence, are excellent candidates for improving abiotic stress tolerance in plants.

Soy G3470 (SEQ ID NO: 2947 and 2948)

G3470 is ortholog of G481 and G482 from *Glycine max*. Among the *Arabidopsis* paralogs in the G481 study group, G3470 is phylogenetically most closely related to G481 itself.

Experimental Observations

The aim of this study was to assess the role of G3470 in drought stress-related tolerance via overexpression, and compare the effects with that of the other G481 orthologs and paralogs.

Twenty 35S::G3470 lines were examined. Half of the transformants exhibited a marked delay in flowering, of about 1 week, and had rather dark narrowed leaves compared to wild-type or non-transformed controls. This same phenotype was noted for G481 and some of its other putative orthologs.

35S::G3470 lines have now been tested in plate based physiology assays; seven of ten lines showed enhanced germination, relative to wild type, when tested in NaCl germination assays. Additionally, two 35S::G3470 lines showed an enhanced performance in a heat growth assay.

Table 48. G3470 35S, Direct promoter-fusion

Line	Germ. in high NaCl	Germ. in high mannitol	Sucrose	ABA	Germ. in heat	Germ. in cold	Growth in heat	Drought	Chilling
301	wt	wt	wt	wt	wt	wt	+	wt	wt
302	+	wt	wt	wt	wt	wt	wt	wt	wt
303	+	wt	wt	wt	wt	wt	+	wt	wt
305	+	wt	wt	wt	wt	wt	wt	wt	wt
306	wt	wt	wt	wt	wt	wt	wt	wt	wt
309	+	wt	wt	wt	wt	wt	wt	wt	wt
310	+	wt	wt	wt	wt	wt	wt	wt	wt
316	+	wt	wt	wt	wt	wt	wt	wt	wt
317	wt	wt	wt	wt	wt	wt	wt	wt	wt
318	+	wt	wt	wt	wt	wt	wt	wt	wt

Potential Applications

The results of the salt germination assay confirm that G481 and its related sequences, including the soy G3470 sequence, are excellent candidates for improving abiotic stress tolerance in plants.

Soy G3471 (SEQ ID NO: 2949 and 2950)

G3471 from *Glycine max* is an ortholog of G481 and G482. Among the *Arabidopsis* paralogs in the G481 study group, G3471 is phylogenetically most closely related to G481.

Experimental Observations

The aim of this study was to assess the role of G3471 in drought stress-related tolerance via overexpression, and compare the effects with that of the other G481 orthologs and paralogs.

Changes in flowering time were seen among the 35S::G3471 lines. A number of lines appeared late flowering, while others showed a marginal acceleration of flowering. Some of the 35S::G3471 lines also showed alterations in leaf shape.

Several lines performed better than controls in abiotic stress assays, including germination in heat, growth in cold conditions, and better drought tolerance in a plate-based assay.

Table 49. G3471 35S, Direct promoter-fusion

Line	Germ. in high NaCl	Germ. in high mannitol	Sucrose	ABA	Germ. in heat	Germ. in cold	Growth in heat	Drought	Chilling
303	wt	wt	wt	wt	+	wt	wt	wt	+
306	wt	wt	wt	wt	wt	wt	wt	wt	wt
307	wt	wt	wt	wt	+	wt	wt	wt	wt
308	wt	wt	wt	wt	wt	wt	wt	+	wt
312	wt	wt	wt	wt	+	wt	wt	wt	wt
328	wt	wt	wt	wt	wt	wt	wt	wt	wt
329	wt	wt	wt	wt	wt	wt	wt	wt	
330	wt	wt	wt	wt	wt	wt	wt	wt	
337	wt	wt	wt	wt	wt	wt	wt	wt	
338	wt	wt	wt	wt	wt	wt	wt	wt	wt

Potential Applications

The results of the salt germination assay confirm that G481 and its related sequences, including the soy G3472 sequence, are excellent candidates for improving abiotic stress tolerance in plants.

Potential Applications

The results of the salt germination assay confirm that G481 and its related sequences, including the soy G3472 sequence, are excellent candidates for improving abiotic stress tolerance in plants.

Soy G3472 (SEQ ID NO: 801 and 2907)

G3472 is an ortholog of G481 and G482 from *Glycine max*, and is a member of the HAP3 subgroup of the CCAAT-box binding transcription factor family. Among the *Arabidopsis* paralogs in the G481 study group, this gene is phylogenetically most closely related to G485.

Experimental Observations

The aim of this study was to assess the role of G3472 in drought stress-related tolerance via overexpression,. 35S::G3472 lines showed no consistent differences in morphology to wild-type or non-transformed controls. G3472 did not produce accelerated flowering in the same manner as did other G485 related genes such as G3474, G3475 and G3476.

Three 35S::G3472 lines were more salt tolerant than wild-type or non-transformed controls in a plate-based assay.

Table 50. G3472 35S, Direct promoter-fusion

Line	Germ. in high NaCl	Germ. in high mannitol	Sucrose	ABA	Germ. in heat	Germ. in cold	Growth in heat	Drought	Chilling
303	wt	wt	wt	wt	wt	wt	wt	wt	wt
304	wt	wt	wt	wt	wt	wt	wt	wt	wt
305	wt	wt	wt	wt	wt	wt	wt	wt	wt
306	wt	wt	wt	wt	wt	wt	wt	wt	wt
307	wt	wt	wt	wt	wt	wt	wt	wt	wt
308	+	wt	wt	wt	wt	wt	wt	wt	wt
311	+	wt	wt	wt	wt	wt	wt	wt	wt
313	wt	wt	wt	wt	wt	wt	wt	wt	wt
314	wt	wt	wt	wt	wt	wt	wt	wt	wt
318	+	wt	wt	wt	wt	wt	wt	wt	-

Potential Applications

The results of the salt germination assay confirm that G481 and its related sequences, including the soy G3472 sequence, are excellent candidates for improving abiotic stress tolerance in plants.

Soy G3475 (SEQ ID NO: 800 and 2908)

G3475 is from *Glycine max* and is a putative ortholog of G481, and is a member of the HAP3 subgroup of the CCAAT-box binding transcription factor family. Among the *Arabidopsis* paralogs in the G481 study group, this gene is phylogenetically most closely related to G485.

Experimental Observations

35S::G3475 lines showed accelerated flowering by about one to two weeks compared to wild-type or non-transformed controls. This same phenotype was also noted for the most closely related *Arabidopsis* gene, G485. Many of these early flowering lines also were smaller than controls.

Growth of four 35S::G3475 lines was more tolerant to cold conditions than wild-type or non-transformed controls.

Table 51. G3475 35S, Direct promoter-fusion

Line	Germ. in high NaCl	Germ. in high mannitol	Sucrose	ABA	Germ. in heat	Germ. in cold	Growth in heat	Drought	Chilling
301	wt	wt	wt	wt	wt	wt	wt	wt	wt
302	wt	wt	wt	wt	wt	wt	wt	wt	+
303	wt	wt	wt	wt	wt	wt	wt	wt	wt
304	wt	wt	wt	wt	wt	wt	wt	wt	+
306	wt	wt	wt	wt	wt	wt	wt	wt	+
307	wt	wt	wt	wt	wt	wt	wt	wt	+
308	wt	wt	wt	wt	wt	wt	wt	wt	wt
309	wt	wt	wt	wt	wt	wt	wt	wt	wt
310	wt	wt	wt	wt	wt	wt	wt	wt	wt
311	wt	wt	wt	wt	wt	wt	wt	wt	wt

Potential Applications

The results of the chilling assay confirm that G481 and its related sequences, including the soy G3475 sequence, are excellent candidates for improving abiotic stress tolerance in plants.

Example XIV: Transformation of Cereal Plants with an Expression Vector

Cereal plants such as, but not limited to, corn, wheat, rice, sorghum, or barley, may also be transformed with the present polynucleotide sequences in pMEN20 or pMEN65 expression vectors for the purpose of modifying plant traits. For example, pMEN020 may be modified to replace the NptII coding region with the BAR gene of *Streptomyces hygroscopicus* that confers resistance to phosphinothricin. The KpnI and BglII sites of the Bar gene are removed by site-directed mutagenesis with silent codon changes.

The cloning vector may be introduced into a variety of cereal plants by means well known in the art such as, for example, direct DNA transfer or *Agrobacterium tumefaciens*-mediated transformation. It is now routine to produce transgenic plants of most cereal crops (Vasil (1994) *Plant Mol. Biol.* 25: 925-937) such as corn, wheat, rice, sorghum (Cassas et al. (1993) *Proc. Natl. Acad. Sci.* 90: 11212-11216, and barley (Wan and Lemeaux (1994) *Plant Physiol.* 104:37-48. DNA transfer methods such as the

microprojectile can be used for corn (Fromm et al. (1990) *Bio/Technol.* 8: 833-839); Gordon-Kamm et al. (1990) *Plant Cell* 2: 603-618; Ishida (1990) *Nature Biotechnol.* 14:745-750), wheat (Vasil et al. (1992) *Bio/Technol.* 10:667-674; Vasil et al. (1993) *Bio/Technol.* 11:1553-1558; Weeks et al. (1993) *Plant Physiol.* 102:1077-1084), rice (Christou (1991) *Bio/Technol.* 9:957-962; Hiei et al. (1994) *Plant J.* 6:271-282; Aldemita and Hodges (1996) *Planta* 199:612-617; and Hiei et al. (1997) *Plant Mol. Biol.* 35:205-218). For most cereal plants, embryogenic cells derived from immature scutellum tissues are the preferred cellular targets for transformation (Hiei et al. (1997) *Plant Mol. Biol.* 35:205-218; Vasil (1994) *Plant Mol. Biol.* 25: 925-937).

Vectors according to the present invention may be transformed into corn embryogenic cells derived from immature scutellar tissue by using microprojectile bombardment, with the A188XB73 genotype as the preferred genotype (Fromm et al. (1990) *Bio/Technol.* 8: 833-839; Gordon-Kamm et al. (1990) *Plant Cell* 2: 603-618). After microprojectile bombardment the tissues are selected on phosphinothricin to identify the transgenic embryogenic cells (Gordon-Kamm et al. (1990) *Plant Cell* 2: 603-618). Transgenic plants are regenerated by standard corn regeneration techniques (Fromm et al. (1990) *Bio/Technol.* 8: 833-839; Gordon-Kamm et al. (1990) *Plant Cell* 2: 603-618).

The plasmids prepared as described above can also be used to produce transgenic wheat and rice plants (Christou (1991) *Bio/Technol.* 9:957-962; Hiei et al. (1994) *Plant J.* 6:271-282; Aldemita and Hodges (1996) *Planta* 199:612-617; and Hiei et al. (1997) *Plant Mol. Biol.* 35:205-218) that coordinately express genes of interest by following standard transformation protocols known to those skilled in the art for rice and wheat (Vasil et al. (1992) *Bio/Technol.* 10:667-674; Vasil et al. (1993) *Bio/Technol.* 11:1553-1558; and Weeks et al. (1993) *Plant Physiol.* 102:1077-1084), where the bar gene is used as the selectable marker.

Example XV: Transformation of Canola with a Plasmid Containing CBF1, CBF2, or CBF3

After identifying homologous genes to CBF1, canola was transformed with a plasmid containing the *Arabidopsis* CBF1, CBF2, or CBF3 genes cloned into the vector pGA643 (An (1987) *Methods Enzymol.* 253: 292). In these constructs the CBF genes were expressed constitutively under the CaMV 35S promoter. In addition, the CBF1 gene was cloned under the control of the *Arabidopsis* COR15 promoter in the same vector pGA643. Each construct was transformed into *Agrobacterium* strain GV3101. Transformed Agrobacteria were grown for 2 days in minimal AB medium containing appropriate antibiotics.

Spring canola (*B. napus* cv. Westar) was transformed using the protocol of Moloney et al. (1989) *Plant Cell Reports* 8: 238, with some modifications as described. Briefly, seeds were sterilized and plated on half strength MS medium, containing 1% sucrose. Plates were incubated at 24° C under 60-80 $\mu E/m^2s$ light using a16 hour light/ 8 hour dark photoperiod. Cotyledons from 4-5 day old seedlings were collected, the petioles cut and dipped into the *Agrobacterium* solution. The dipped cotyledons were placed on co-cultivation medium at a density of 20 cotyledons/plate and incubated as described above for

3 days. Explants were transferred to the same media, but containing 300 mg/l timentin (SmithKline Beecham, PA) and thinned to 10 cotyledons/plate. After 7 days explants were transferred to Selection/Regeneration medium. Transfers were continued every 2-3 weeks (2 or 3 times) until shoots had developed. Shoots were transferred to Shoot-Elongation medium every 2-3 weeks. Healthy looking shoots were transferred to rooting medium. Once good roots had developed, the plants were placed into moist potting soil.

The transformed plants were then analyzed for the presence of the NPTII gene/ kanamycin resistance by ELISA, using the ELISA NPTII kit from 5Prime-3Prime Inc. (Boulder, CO). Approximately 70% of the screened plants were NPTII positive. Only those plants were further analyzed.

From Northern blot analysis of the plants that were transformed with the constitutively expressing constructs, showed expression of the CBF genes and all CBF genes were capable of inducing the *Brassica napus* cold-regulated gene BN115 (homolog of the *Arabidopsis* COR15 gene). Most of the transgenic plants appear to exhibit a normal growth phenotype. As expected, the transgenic plants are more freezing tolerant than the wild-type plants. Using the electrolyte leakage of leaves test, the control showed a 50% leakage at –2 to –3° C. Spring canola transformed with either CBF1 or CBF2 showed a 50% leakage at –6 to –7° C. Spring canola transformed with CBF3 shows a 50% leakage at about –10 to –15° C. Winter canola transformed with CBF3 may show a 50% leakage at about –16 to –20° C. Furthermore, if the spring or winter canola are cold acclimated the transformed plants may exhibit a further increase in freezing tolerance of at least –2° C.

To test salinity tolerance of the transformed plants, plants were watered with 150 mM NaCl. Plants overexpressing CBF1, CBF2 or CBF3 grew better compared with plants that had not been transformed with CBF1, CBF2 or CBF3.

These results demonstrate that equivalogs of *Arabidopsis* transcription factors can be identified and shown to confer similar functions in non-*Arabidopsis* plant species.

Example XVI: Cloning of transcription factor promoters

Promoters are isolated from transcription factor genes that have gene expression patterns useful for a range of applications, as determined by methods well known in the art (including transcript profile analysis with cDNA or oligonucleotide microarrays, Northern blot analysis, semi-quantitative or quantitative RT-PCR). Interesting gene expression profiles are revealed by determining transcript abundance for a selected transcription factor gene after exposure of plants to a range of different experimental conditions, and in a range of different tissue or organ types, or developmental stages. Experimental conditions to which plants are exposed for this purpose includes cold, heat, drought, osmotic challenge, varied hormone concentrations (ABA, GA, auxin, cytokinin, salicylic acid, brassinosteroid), pathogen and pest challenge. The tissue types and developmental stages include stem,

root, flower, rosette leaves, cauline leaves, siliques, germinating seed, and meristematic tissue. The set of expression levels provides a pattern that is determined by the regulatory elements of the gene promoter.

Transcription factor promoters for the genes disclosed herein are obtained by cloning 1.5 kb to 2.0 kb of genomic sequence immediately upstream of the translation start codon for the coding sequence of the encoded transcription factor protein. This region includes the 5'-UTR of the transcription factor gene, which can comprise regulatory elements. The 1.5 kb to 2.0 kb region is cloned through PCR methods, using primers that include one in the 3' direction located at the translation start codon (including appropriate adaptor sequence), and one in the 5' direction located from 1.5 kb to 2.0 kb upstream of the translation start codon (including appropriate adaptor sequence). The desired fragments are PCR-amplified from *Arabidopsis* Col-0 genomic DNA using high-fidelity Taq DNA polymerase to minimize the incorporation of point mutation(s). The cloning primers incorporate two rare restriction sites, such as NotI and SfiI, found at low frequency throughout the *Arabidopsis* genome. Additional restriction sites are used in the instances where a NotI or SfiI restriction site is present within the promoter.

The 1.5-2.0 kb fragment upstream from the translation start codon, including the 5'-untranslated region of the transcription factor, is cloned in a binary transformation vector immediately upstream of a suitable reporter gene, or a transactivator gene that is capable of programming expression of a reporter gene in a second gene construct. Reporter genes used include green fluorescent protein (and related fluorescent protein color variants), beta-glucuronidase, and luciferase. Suitable transactivator genes include LexA-GAL4, along with a transactivatable reporter in a second binary plasmid (as disclosed in U.S. patent application 09/958,131, incorporated herein by reference). The binary plasmid(s) is transferred into *Agrobacterium* and the structure of the plasmid confirmed by PCR. These strains are introduced into *Arabidopsis* plants as described in other examples, and gene expression patterns determined according to standard methods know to one skilled in the art for monitoring GFP fluorescence, beta-glucuronidase activity, or luminescence.

All publications and patent applications mentioned in this specification are herein incorporated by reference to the same extent as if each individual publication or patent application was specifically and individually indicated to be incorporated by reference.

The present invention is not limited by the specific embodiments described herein. The invention now being fully described, it will be apparent to one of ordinary skill in the art that many changes and modifications can be made thereto without departing from the spirit or scope of the appended claims. Modifications that become apparent from the foregoing description and accompanying figures fall within the scope of the claims.

What is claimed is:

1. A transgenic plant having increased abiotic stress tolerance as compared to non-transgenic plants of the same species;

wherein the transgenic plant comprises in its genome a transgene encoding a polypeptide member of the G482 subclade of the non-LEC1-like clade of proteins of the L1L-related CCAAT transcription factor family, wherein overexpression of the polypeptide member confers said abiotic stress tolerance.

2. The transgenic plant of Claim 1, wherein the polypeptide member comprises a B domain;

wherein the B domain is sufficiently homologous to the B domain of SEQ ID NO: 88 that the B domain binds to DNA at a transcription regulating region comprising the motif CCAAT;

wherein said binding regulates transcription of the DNA;

and said regulation of transcription confers increased abiotic stress tolerance in the transgenic plant as compared to non-transgenic plants of the same species.

3. The transgenic plant of Claim 1, wherein the transgene is derived from a monocotyledonous plant.

4. The transgenic plant of Claim 1, wherein the polypeptide member of the CCAAT-box binding transcription factor family is at least 81% identical to a B domain of a polypeptide sequence selected from the group consisting of SEQ ID NO: 88, 790, 794, 796, 798, 800, 801, 805 and 806.

5. The transgenic plant of Claim 1, wherein the polypeptide sequence is selected from the group consisting of SEQ ID NO: 88, 790, 794, 796, 798, 800, 801, 805 and 806.

6. The transgenic plant of Claim 1, wherein the transgene comprises a polynucleotide sequence selected from the group consisting of SEQ ID NO: 87, 2907, 2908, 2909, 2910, 2911, 2912, 2913 and 2914.

7. The transgenic plant of Claim 1, wherein said abiotic stress tolerance is selected from the group consisting of heat tolerance, drought stress tolerance, cold tolerance, and salt stress tolerance.

8. The transgenic plant of Claim 1, wherein the plant is selected from the group consisting of: soybean, wheat, corn, potato, cotton, rice, oilseed rape, sunflower, alfalfa, clover, sugarcane, turf, banana, blackberry, blueberry, strawberry, raspberry, cantaloupe, carrot, cauliflower, coffee, cucumber, eggplant,

grapes, honeydew, lettuce, mango, melon, onion, papaya, peas, peppers, pineapple, pumpkin, spinach, squash, sweet corn, tobacco, tomato, watermelon, mint and other labiates, fruit trees, rosaceous fruits, citrus, and brassicas.

9. The transgenic plant of Claim 1, further comprising a constitutive, inducible, or tissue-specific promoter operably linked to said recombinant polynucleotide.

10. The transgenic plant of Claim 1, wherein the recombinant polynucleotide is incorporated into an expression vector comprising one or more regulatory elements that are effective to control expression of said recombinant polynucleotide in a target plant

11. The transgenic plant of Claim 1, wherein the transgenic plant is a cultured host cell.

12. Seed from the transgenic plant according to Claim 1

13. A method for producing a transgenic plant having increased tolerance to abiotic stress, the method steps comprising:

(a) providing an expression vector comprising

(i) a polynucleotide sequence comprising a nucleotide sequence encoding a polypeptide that comprises a member of G482 subclade of the non-LEC1-like clade of proteins of the L1L-related CCAAT transcription factor family; wherein

said polypeptide comprises a B domain, and the B domain is sufficiently homologous to the B domain of SEQ ID NO: 88 that the B domain binds to DNA at a transcription regulating region comprising the motif CCAAT;

and wherein said binding regulates transcription of the DNA, and said regulation of transcription confers increased abiotic stress tolerance in the transgenic plant as compared to non-transgenic plants of the same species; and

(ii) regulatory elements operably linked to the nucleotide sequence, said regulatory elements being effective to control expression of said nucleotide sequence in a target plant;

(b) introducing the expression vector into a plant cell;

(c) growing the plant cell into a plant and allowing the plant to overexpress the polypeptide; and;

(d) identifying one or more abiotic stress tolerant plants so produced by comparing said one or more abiotic stress tolerant plants with one or more non-transgenic plants of the same species.

14. The method of Claim 13, wherein said abiotic stress tolerance is selected from the group consisting of heat tolerance, cold tolerance, drought stress tolerance and salt stress tolerance.

15. The method of Claim 13, the method steps further comprising:

(e) crossing one of said abiotic stress tolerant plants with itself or another plant;

(f) selecting seed that develops as a result of said crossing; and growing a progeny plant from the seed, thus producing a transgenic progeny plant having increased tolerance to abiotic stress as compared to non-transgenic plants of the same species.

16. The method of Claim 15, wherein:

said progeny plant expresses mRNA that encodes a DNA-binding protein that binds to a CCAAT DNA regulatory sequence and induces expression of a plant trait gene; and

said mRNA is expressed at a level greater than a non-transformed plant that does not overexpress said DNA-binding protein.

17. A method for increasing a plant's tolerance to abiotic stress, said method comprising:

(a) providing a vector comprising:

(i) a polynucleotide sequence comprising a nucleotide sequence encoding a polypeptide that comprises a member of G482 subclade of the non-LEC1-like clade of proteins of the L1L-related CCAAT transcription factor family; wherein

said polypeptide comprises a B domain, and the B domain is sufficiently homologous to the B domain of SEQ ID NO: 88 that the B domain binds to DNA at a transcription regulating region comprising the motif CCAAT;

and wherein said binding regulates transcription of the DNA, and said regulation of transcription confers increased abiotic stress tolerance in the transgenic plant as compared to non-transgenic plants of the same species; and

(ii) regulatory elements operably linked to the nucleotide sequence, said regulatory elements being effective to control expression of said nucleotide sequence in a target plant; and

(b) transforming the target plant with the vector to generate a transformed plant with increased tolerance to abiotic stress compared to non-transgenic plants of the same species.

18. The method of Claim 17, wherein the polynucleotide is selected from the group consisting of:

(i) SEQ ID NO: 147, 2907, 2908, 2909, 2910, 2911, 2912, 2913 and 2914; and

(ii) a nucleotide sequence encoding a polypeptide that comprises a B domain that is at least 81% identical with the B domain of SEQ ID NO: 88.

19. The method of Claim 17, wherein said abiotic stress tolerance is selected from the group consisting of heat tolerance, cold germination, drought stress tolerance and salt stress tolerance.

20. A transgenic plant having increased abiotic stress tolerance as compared to non-transgenic plants of the same species;

wherein the transgenic plant comprises in its genome a transgene encoding a polypeptide member of the MYB-related transcription factor family, wherein overexpression of the polypeptide member confers said abiotic stress tolerance;

and wherein the polypeptide member of the MYB-related transcription factor family is derived from a plant other than *Arabidopsis*, and the polypeptide member is from the G682 subclade of MYB-related transcription factors.

21. The transgenic plant of Claim 20, wherein the polypeptide member of the MYB-related transcription factor family comprises a MYB-related domain;

wherein the MYB-related domain is sufficiently homologous to the MYB-related domain of SEQ ID NO: 148 that the MYB-related domain binds to DNA at a transcription regulating region;

wherein said binding regulates transcription of the DNA;

and said regulation of transcription confers increased abiotic stress tolerance in the transgenic plant as compared to non-transgenic plants of the same species.

22. The transgenic plant of Claim 20, wherein the transgene is derived from a monocotyledonous plant.

23. The transgenic plant of Claim 20, wherein the polypeptide member of the MYB-related transcription factor family is at least 60% identical to a MYB-related domain of a polypeptide sequence selected from the group consisting of SEQ ID NO: 559, 1082, 1083, 1084, 1086, 1087, 1088 and 1089.

24. The transgenic plant of Claim 20, wherein the polypeptide sequence is selected from the group consisting of SEQ ID NO: 148, 559, 1082, 1083, 1084, 1086, 1087, 1088 and 1089.

25. The transgenic plant of Claim 20, wherein the transgene comprises a polynucleotide sequence that hybridizes under stringent conditions to the complement of a member of the group selected from SEQ ID NO: 147, 2915, 2916, 2917, 2918, 2919, 2920, 2921, and 2922.

26. The transgenic plant of Claim 20, wherein the transgene comprises a polynucleotide sequence selected from the group consisting of 147, 2915, 2916, 2917, 2918, 2919, 2920, 2921, and 2922.

27. The transgenic plant of Claim 20, wherein said abiotic stress tolerance is selected from the group consisting of heat tolerance, drought stress tolerance, cold tolerance, and salt stress tolerance.

28. The transgenic plant of Claim 20, wherein the transgenic plant is characterized by altered sugar sensing as compared to non-transgenic plants of the same species.

29. The transgenic plant of Claim 20, wherein the plant is selected from the group consisting of: soybean, wheat, corn, potato, cotton, rice, oilseed rape, sunflower, alfalfa, clover, sugarcane, turf, banana, blackberry, blueberry, strawberry, raspberry, cantaloupe, carrot, cauliflower, coffee, cucumber, eggplant, grapes, honeydew, lettuce, mango, melon, onion, papaya, peas, peppers, pineapple, pumpkin, spinach, squash, sweet corn, tobacco, tomato, watermelon, mint and other labiates, fruit trees, rosaceous fruits, citrus, and brassicas.

30 The transgenic plant of Claim 20, further comprising a constitutive, inducible, or tissue-specific promoter operably linked to said recombinant polynucleotide.

31. The transgenic plant of Claim 20, wherein the recombinant polynucleotide is incorporated into an expression vector comprising one or more regulatory elements that are effective to control expression of said recombinant polynucleotide in a target plant

32. The transgenic plant of Claim 20, wherein the transgenic plant is a cultured host cell.

33. Seed from the transgenic plant according to Claim 20.

34. A method for producing a transgenic plant having increased tolerance to abiotic stress, the method steps comprising:

(a) providing an expression vector comprising

(i) a polynucleotide sequence comprising a nucleotide sequence encoding a polypeptide that comprises a polypeptide member of the MYB-related transcription factor family; wherein said polypeptide comprises a MYB-related domain, and the MYB-related domain is sufficiently homologous to the MYB-related domain of SEQ ID NO: 148 that the MYB-related domain binds to DNA at a transcription regulating region;

and wherein said binding regulates transcription of the DNA, and said regulation of transcription confers increased abiotic stress tolerance in the transgenic plant as compared to non-transgenic plants of the same species; and

(ii) regulatory elements operably linked to the nucleotide sequence, said regulatory elements being effective to control expression of said nucleotide sequence in a target plant;

(b) introducing the expression vector into a plant cell;

(c) growing the plant cell into a plant and allowing the plant to overexpress the polypeptide; and;

(d) identifying one or more abiotic stress tolerant plants so produced by comparing said one or more abiotic stress tolerant plants with one or more non-transgenic plants of the same species.

35. The method of Claim 34, wherein said abiotic stress tolerance is selected from the group consisting of heat tolerance, cold tolerance, drought stress tolerance and salt stress tolerance.

36. The method of Claim 34, the method steps further comprising:

(e) crossing one of said abiotic stress tolerant plants with itself or another plant;

(f) selecting seed that develops as a result of said crossing; and growing a progeny plant from the seed, thus producing a transgenic progeny plant having increased tolerance to abiotic stress as compared to non-transgenic plants of the same species.

37. The method of Claim 36, wherein:

said progeny plant expresses mRNA that encodes a DNA-binding protein that binds to a DNA regulatory sequence and induces expression of a plant trait gene; and

said mRNA is expressed at a level greater than a non-transformed plant that does not overexpress said DNA-binding protein.

38. A method for increasing a plant's tolerance to abiotic stress, said method comprising:

(a) providing a vector comprising:

(i) a polynucleotide sequence comprising a nucleotide sequence encoding a polypeptide that comprises a member of G682 subclade of MYB-related transcription factors; wherein said polypeptide comprises a MYB-related domain, and the MYB-related domain is sufficiently homologous to the MYB-related domain of SEQ ID NO: 148 that the MYB-related domain binds to DNA at a transcription regulating region;

and wherein said binding regulates transcription of the DNA, and said regulation of transcription confers increased abiotic stress tolerance in the transgenic plant as compared to non-transgenic plants of the same species; and

(ii) regulatory elements operably linked to the nucleotide sequence, said regulatory elements being effective to control expression of said nucleotide sequence in a target plant; and

(b) transforming the target plant with the vector to generate a transformed plant with increased tolerance to abiotic stress compared to non-transgenic plants of the same species.

39. The method of Claim 38, wherein the polynucleotide is selected from the group consisting of:

(i) SEQ ID NO: 147, 2907, 2915, 2916, 2917, 2918, 2919, 2920, 2921, and 2922;

(ii) a nucleotide sequence that hybridizes to the nucleotide sequence of (i) under stringent conditions that include two wash steps of 0.1 x SSC to 2.0 x SSC and 0.1% SDS at 50-65°C, for 10 - 30 min per step; and

(iii) a nucleotide sequence encoding a polypeptide that comprises a MYB-related domain that is at least 60% identical with the MYB-related domain of SEQ ID NO: 148.

40. The method of Claim 39, wherein the stringent conditions include two wash steps of 0.1 x SSC and 0.1% SDS at 65°C, for 10 - 30 min per step.

41. The method of Claim 38, wherein said abiotic stress tolerance is selected from the group consisting of heat tolerance, cold germination, drought stress tolerance and salt stress tolerance.

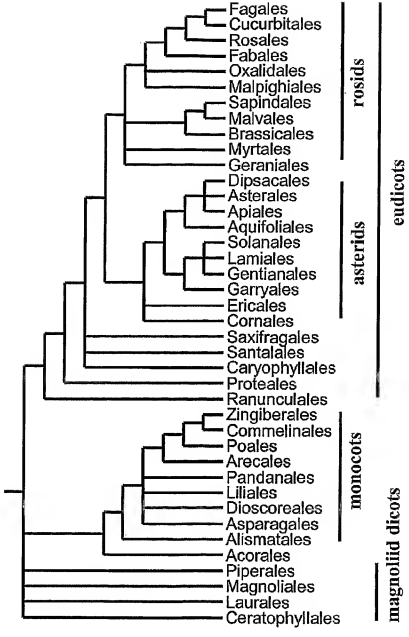

Fagales
Cucurbitales
Rosales
Fabales
Oxalidales
Malpighiales
Sapindales
Malvales
Brassicales
Myrtales
Geraniales
Dipsacales
Asterales
Apiales
Aquifoliales
Solanales
Lamiales
Gentianales
Garryales
Ericales
Cornales
Saxifragales
Santalales
Caryophyllales
Proteales
Ranunculales
Zingiberales
Commelinales
Poales
Arecales
Pandanales
Liliales
Dioscoreales
Asparagales
Alismatales
Acorales
Piperales
Magnoliales
Laurales
Ceratophyllales
rosids
asterids
eudicots
monocots
magnoliid dicots

FIGURE 1

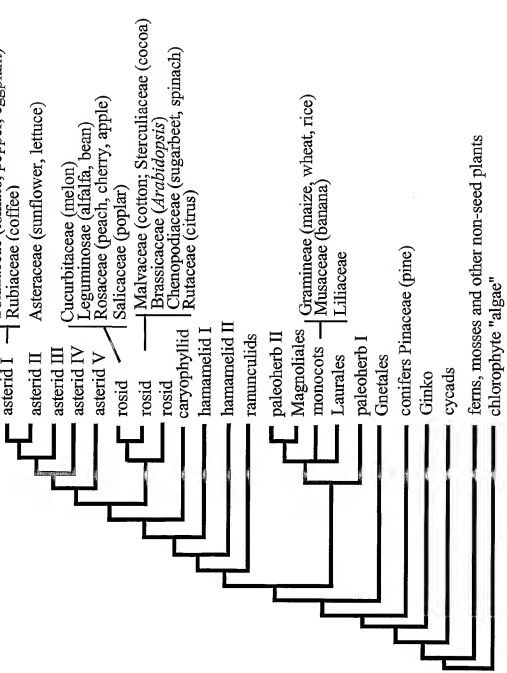

asterid I
asterid II
asterid III
asterid IV
asterid V
rosid
rosid
rosid
caryophyllid
hamamelid I
hamamelid II
ranunculids
paleoherb II
Magnoliales
monocots
Laurales
paleoherb I
Gnetales
conifers Pinaceae (pine)
Ginko
cycads
ferns, mosses and other non-seed plants
chlorophyte "algae"
Solanaceae (tomato, pepper, eggplant)
Rubiaceae (coffee)
Asteraceae (sunflower, lettuce)
Cucurbitaceae (melon)
Leguminosae (alfalfa, bean)
Rosaceae (peach, cherry, apple)
Salicaceae (poplar)
Malvaceae (cotton; Sterculiaceae (cocoa)
Brassicaceae (Arabidopsis)
Chenopodiaceae (sugarbeet, spinach)
Rutaceae (citrus)
Gramineae (maize, wheat, rice)
Musaceae (banana)
Liliaceae

FIGURE 2


SEQIDNO148
SEQIDNO1972
SEQIDNO38
SEQIDNO2142
SEQIDNO1084
SEQIDNO1085
SEQIDNO1086
SEQIDNO1083
SEQIDNO1087
SEQIDNO1088
SEQIDNO559
SEQIDNO1082
SEQIDNO1081
SEQIDNO1089
SEQIDNO1090

FIGURE 3A

	100	110	120
SEQ ID NO 148			
SEQ ID NO 1972	R R D F F R K		
SEQ ID NO 38	R R R L N N S P F F S T S P L N L Q E N L K L		
SEQ ID NO 2142	R R Q L H S S S H K H T K P H R P R F S I Y P S		
SEQ ID NO 1084	R - - N - G S K T Q D S		
SEQ ID NO 1085			
SEQ ID NO 1086			
SEQ ID NO 1083			
SEQ ID NO 1087			
SEQ ID NO 1088			
SEQ ID NO 559			
SEQ ID NO 1082			
SEQ ID NO 1081	M		
SEQ ID NO 1089			
SEQ ID NO 1090			

FIGURE 3B

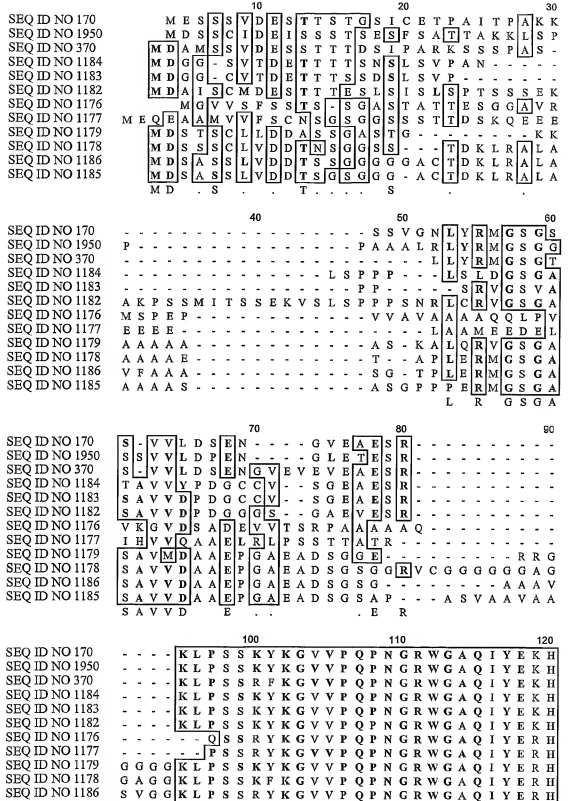

10 20 30
SEQ ID NO 170 M E S S V D E S T T S T G S I C E T P A I T P A K K
SEQ ID NO 1950 M D S S C I D E I S S S T S E S F S A T T A K K L S P
SEQ ID NO 370 M D A M S S V D E S S T T T D S I P A R K S S S P A S -
SEQ ID NO 1184 M D G G - S V T D E T T T T S N S L S V P A N - - - - -
SEQ ID NO 1183 M D G G - C V T D E T T T S S D S L S V P - - - - - - -
SEQ ID NO 1182 M D A I S C M D E S T T T E S L S I S L S P T S S S E K
SEQ ID NO 1176 M G V V S F S S T S - S G A S T A T T E S G G A V R
SEQ ID NO 1177 M E Q E A A M V V F S C N S G S G G S S S T T D S K Q E E E
SEQ ID NO 1179 M D S T S C L L D D A S S G A S T G - - - - - - - - K K
SEQ ID NO 1178 M D S S S C L V D D T N S G G S S - - - T D K L R A L A
SEQ ID NO 1186 M D S A S S L V D D T S S G G G G A C T D K L R A L A
SEQ ID NO 1185 M D S A S S L V D D T S G S G G G - A C T D K L R A L A
M D . S . . T . . . . S . .
40 50 60
SEQ ID NO 170 - - - - - - - - - - - - - - - - S S V G N L Y R M G S G S
SEQ ID NO 1950 P - - - - - - - - - - - - - - - - P A A A L R L Y R M G S G G
SEQ ID NO 370 - L L Y R M G S G T
SEQ ID NO 1184 - - - - - - - - - - - - - - L S P P P - - - L S L D G S G A
SEQ ID NO 1183 - - - - - - - - - - - - - - P P - - - - - S R V G S V A
SEQ ID NO 1182 A K P S S M I T S S E K V S L S P P P S N R L C R V G S G A
SEQ ID NO 1176 M S P E P - - - - - - - - - - - - V V A V A A A A Q Q L P V
SEQ ID NO 1177 E E E E - - - - - - - - - - - - - - L A A M E E D E L
SEQ ID NO 1179 A A A A A - - - - - - - - - - - - A S - K A L Q R V G S G A
SEQ ID NO 1178 A A A A E - - - - - - - - - - - - T - - A P L E R M G S G A
SEQ ID NO 1186 V F A A A - - - - - - - - - - - - S G - T P L E R M G S G A
SEQ ID NO 1185 A A A A S - - - - - - - - - - - - A S G P P F E R M G S G A
L R G S G A
70 80 90
SEQ ID NO 170 S - V V L D S E N - - - - G V E A E S R - - - - - - - - - -
SEQ ID NO 1950 S S V V L D P E N - - - - G L E T E S R - - - - - - - - - -
SEQ ID NO 370 S - V V L D S E N G V E V E V E A E S R - - - - - - - - - -
SEQ ID NO 1184 T A V V Y P D G C C V - - S G E A E S R - - - - - - - - - -
SEQ ID NO 1183 S A V V D P D G C C V - - S G E A E S R - - - - - - - - - -
SEQ ID NO 1182 S A V V D P D G G G S - - G A E V E S R - - - - - - - - - -
SEQ ID NO 1176 V K G V D S A D E V V T S R P A A A A Q - - - - - - - - - -
SEQ ID NO 1177 I H V V Q A A E L R L P S S T T A T R - - - - - - - - - - -
SEQ ID NO 1179 S A V M D A A E P G A E A D S G E - - - - - - - - - - R R G
SEQ ID NO 1178 S A V V D A A E P G A E A D S G S G G R V C G G G G G G A G
SEQ ID NO 1186 S A V V D A A E P G A E A D S G S G - - - - - - - - A A A V
SEQ ID NO 1185 S A V V D A A E P G A E A D S G S A P - - - A S V A A V A A
S A V V D E . . . E R
100 110 120
SEQ ID NO 170 - - - - K L P S S K Y K G V V P Q P N G R W G A Q I Y E K H
SEQ ID NO 1950 - - - - K L P S S K Y K G V V P Q P N G R W G A Q I Y E K H
SEQ ID NO 370 - - - - K L P S S R F K G V V P Q P N G R W G A Q I Y E K H
SEQ ID NO 1184 - - - - K L P S S K Y K G V V P Q P N G R W G A Q I Y E K H
SEQ ID NO 1183 - - - - K L P S S K Y K G V V P Q P N G R W G A Q I Y E K H
SEQ ID NO 1182 - - - - K L P S S K Y K G V V P Q P N G R W G A Q I Y E K H
SEQ ID NO 1176 - - - - - Q S S R Y K G V V P Q P N G R W G A Q I Y E R H
SEQ ID NO 1177 - - - - - P S S R Y K G V V P Q P N G R W G A Q I Y E R H
SEQ ID NO 1179 G G G G K L P S S K Y K G V V P Q P N G R W G A Q I Y E R H
SEQ ID NO 1178 G A G G K L P S S K F K G V V P Q P N G R W G A Q I Y E R H
SEQ ID NO 1186 S V G G K L P S S R Y K G V V P Q P N G R W G A Q I Y E R H
SEQ ID NO 1185 G V G G K L P S S R Y K G V V P Q P N G R W G A Q I Y E R H
K L P S S K Y K G V V P Q P N G R W G A Q I Y E . H

FIGURE 4A

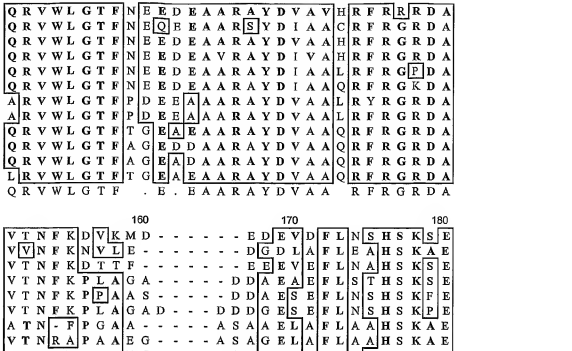

130 140 150
SEQ ID NO 170 QRVWLGTFNEEDEAARAYDVAVHRFRRRDA
SEQ ID NO 1950 QRVWLGTFNEQEEAARSYDIAACRFRGRDA
SEQ ID NO 370 QRVWLGTFNEEDEAARAYDVAAHRFRGRDA
SEQ ID NO 1184 QRVWLGTFNEEDEAVRAYDIVAHRFRGRDA
SEQ ID NO 1183 QRVWLGTFNEEDEAARAYDIAALRFRGPDA
SEQ ID NO 1182 QRVWLGTFNEEDEAARAYDIAAQRFRGKDA
SEQ ID NO 1176 ARVWLGTFPDEEEAARAYDVAALRYRGRDA
SEQ ID NO 1177 ARVWLGTFPDEEEAARAYDVAALRFRGRDA
SEQ ID NO 1179 QRVWLGTFTGEAEAARAYDVAAQRFRGRDA
SEQ ID NO 1178 QRVWLGTFAGEDDAARAYDVAAQRFRGRDA
SEQ ID NO 1186 QRVWLGTFAGEADAARAYDVAAQRFRGRDA
SEQ ID NO 1185 LRVWLGTFTGEAEAARAYDVAAQRFRGRDA
QRVWLGTF.E.EAARAYDVAA RFRGRDA
160 170 180
SEQ ID NO 170 VTNFKDVKMD-------EDEVDFLNSHSKSE
SEQ ID NO 1950 VVNFKNVLE-------DGDLAFLEAHSKAE
SEQ ID NO 370 VTNFKDTTF-------EEEVEFLNAHSKSE
SEQ ID NO 1184 VTNFKPLAGA-----DDAEAEFLSTHSKSE
SEQ ID NO 1183 VTNFKPPAAS-----DDAESEFLNSHSKFE
SEQ ID NO 1182 VTNFKPLAGAD---DDDGESEFLNSHSKPE
SEQ ID NO 1176 ATN-FPGAA-----ASAAELAFLAAHSKAE
SEQ ID NO 1177 VTNRAPAAEG-----ASAGELAFLAAHSKAE
SEQ ID NO 1179 VTNFRPLAES--DPEAAVELRFLASRSKAE
SEQ ID NO 1178 VTNFRPLAEA--DPDAAAELRFLATRSKAE
SEQ ID NO 1186 VTNFRPLADA--DPDAAAELRFLASRSKAE
SEQ ID NO 1185 VTNFRPLAESDLDPDAAAELRFLASRSKAE
VTNF.P.A .EL FL .HSKAE
190 200 210
SEQ ID NO 170 IVDMLRKHTYNEELEQSKRR-----RNGNG
SEQ ID NO 1950 IVDMLRKHTYADELEQNNKRQLFLSVDANG
SEQ ID NO 370 IVDMLRKHTYKEELDQQRKRN-----RDGNG
SEQ ID NO 1184 IVDMLRRHTYDNELQQSTRGG-----RRR--
SEQ ID NO 1183 IVDMLRKHTYDDELQQSTRGG-----RRR--
SEQ ID NO 1182 IVDMLRKHTYNDELEQSKRSRGVV-RRRGS
SEQ ID NO 1176 IVDMLRKHTYADELRQGLRRGR----GMGA
SEQ ID NO 1177 VVDMLRKHTYDDELQQGLRRG-------S--
SEQ ID NO 1179 VVDMLRKHTYLEELTQNKRAFA----AISP
SEQ ID NO 1178 VVDMLRKHTYFDELAQSKRTFA----ASTP
SEQ ID NO 1186 VVDMLRKHTYFDELAQNKRAFA----AASA
SEQ ID NO 1185 VVDMLRKHTYGEELAQNRRAFA----AAAA
IVDMLRKHTY DEL Q .R
220 230 240
SEQ ID NO 170 NMTRTLLTSGLSNDGVSTTGF------RSA
SEQ ID NO 1950 KRNGSSTTQND-----KVL-------KTC
SEQ ID NO 370 KETTAFALASM-----VVMTGF------KTA
SEQ ID NO 1184 --RDAETASSG-----AFDA-------KAR
SEQ ID NO 1183 --LDADTASSG-----VFDA-------KAR
SEQ ID NO 1182 AAAGTANSISG-----ACFT-------KAR
SEQ ID NO 1176 --RAQPTPS-----------------WAR
SEQ ID NO 1177 --RAQPTPR-----------------WAR
SEQ ID NO 1179 P-PPKHPAS------SPTSS------SAR
SEQ ID NO 1178 S-AATTTAS--LSNGHLSSPRSPFAPAAR
SEQ ID NO 1186 A-TASSLANNPPSYASLSPATATAAAAAR
SEQ ID NO 1185 S-LASPQLP----PAKNTSP------AAR
AR

FIGURE 4B

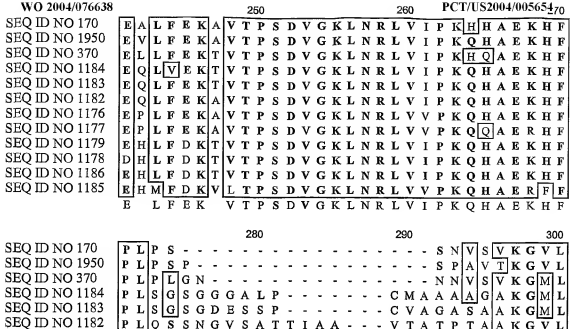

250
260
SEQ ID NO 170
SEQ ID NO 1950
SEQ ID NO 370
SEQ ID NO 1184
SEQ ID NO 1183
SEQ ID NO 1182
SEQ ID NO 1176
SEQ ID NO 1177
SEQ ID NO 1179
SEQ ID NO 1178
SEQ ID NO 1186
SEQ ID NO 1185
280
290
300
310
320
330
340
350
360

FIGURE 4C

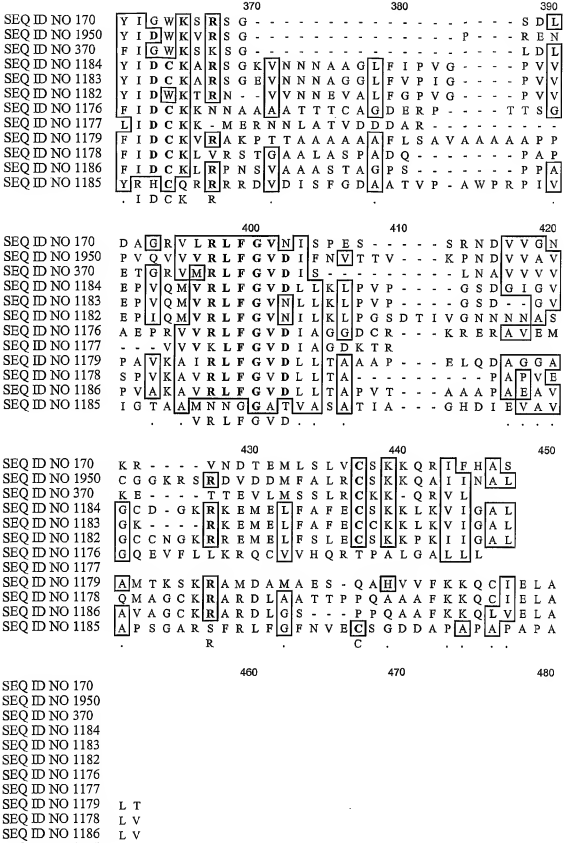
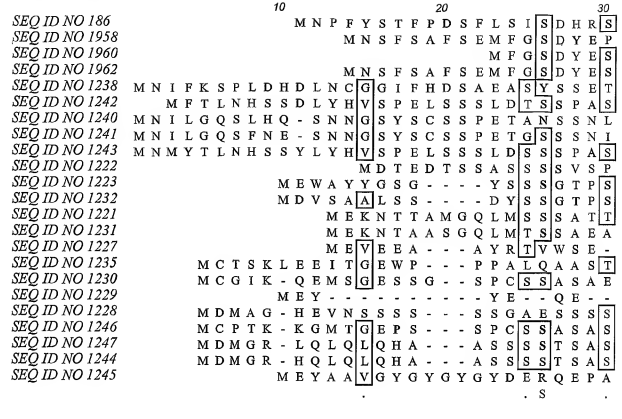

370 380 390
SEQ ID NO 170 Y I G W K S R S G - - - - - - - - - - - - - - - - - - S D L
SEQ ID NO 1950 Y I D W K V R S G - - - - - - - - - - - - - - - - P - - - R E N
SEQ ID NO 370 F I G W K S K S G - - - - - - - - - - - - - - - - - - L D L
SEQ ID NO 1184 Y I D C K A R S G K V N N N A A G L F I P V G - - - - P V V
SEQ ID NO 1183 Y I D C K A R S G E V N N N A G G L F V P I G - - - - P V V
SEQ ID NO 1182 Y I D W K T R N - - V V N N E V A L F G P V G - - - - P V V
SEQ ID NO 1176 F I D C K K N N A A A T T T C A G D E R P - - - - T T S G
SEQ ID NO 1177 L I D C K K - M E R N N L A T V D D D A R - - - - - - - - -
SEQ ID NO 1179 F I D C K V R A K P T T A A A A A F L S A V A A A A A P P
SEQ ID NO 1178 F I D C K L V R S T G A A L A S P A D Q - - - - - - - P A P
SEQ ID NO 1186 F I D C K L R P N S V A A A S T A G P S - - - - - - - P P A
SEQ ID NO 1185 Y R H C Q R R R D V D I S F G D A A T V P - A W P R P I V
. I D C K R . . .
400 410 420
SEQ ID NO 170 D A G R V L R L F G V N I S P E S - - - - - S R N D V V G N
SEQ ID NO 1950 P V Q V V V R L F G V D I F N V T T V - - - K P N D V V A V
SEQ ID NO 370 E T G R V M R L F G V D I S - - - - - - - - - L N A V V V V
SEQ ID NO 1184 E F V Q M V R L F G V D L L K L P V P - - - - G S D G I G V
SEQ ID NO 1183 E F V Q M V R L F G V N L L K L P V P - - - - G S D - - G V
SEQ ID NO 1182 E P I Q M V R L F G V N I L K L P G S D T I V G N N N N A S
SEQ ID NO 1176 A E P R V V R L F G V D I A G G D C R - - - K R E R A V E M
SEQ ID NO 1177 - - - V V V K L F G V D I A G D K T R
SEQ ID NO 1179 P A V K A I R L F G V D L L T A A A P - - - E L Q D A G G A
SEQ ID NO 1178 S P V K A V R L F G V D L L T A - - - - - - - - - P A P V E
SEQ ID NO 1186 P V A K A V R L F G V D L L T A P V T - - - A A A P A E A V
SEQ ID NO 1185 I G T A A M N N G G A T V A S A T I A - - - G H D I E V A V
. . V R L F G V D . . . . . .
430 440 450
SEQ ID NO 170 K R - - - - V N D T E M L S L V C S K K Q R I F H A S
SEQ ID NO 1950 C G G K R S R D V D D M F A L R C S K K Q A I I N A L
SEQ ID NO 370 K E - - - - T T E V L M S S L R C K K - Q R V L
SEQ ID NO 1184 G C D - G K R K E M E L F A F E C S K K L K V I G A L
SEQ ID NO 1183 G K - - - - R K E M E L F A F E C C K K L K V I G A L
SEQ ID NO 1182 G C C N G K R R E M E L F S L E C S K K P K I I G A L
SEQ ID NO 1176 G Q E V F L L K R Q C V V H Q R T P A L G A L L L
SEQ ID NO 1177
SEQ ID NO 1179 A M T K S K R A M D A M A E S - Q A H V V F K K Q C I E L A
SEQ ID NO 1178 Q M A G C K R A R D L A A T T P P Q A A A F K K Q C I E L A
SEQ ID NO 1186 A V A G C K R A R D L G S - - - P P Q A A A F K K Q L V E L A
SEQ ID NO 1185 A P S G A R S F R L F G F N V E C S G D D A P A P A P A P A
. R . C . . . .
460 470 480
SEQ ID NO 170
SEQ ID NO 1950
SEQ ID NO 370
SEQ ID NO 1184
SEQ ID NO 1183
SEQ ID NO 1182
SEQ ID NO 1176
SEQ ID NO 1177
SEQ ID NO 1179 L T
SEQ ID NO 1178 L V
SEQ ID NO 1186 L V
SEQ ID NO 1185 E V E Y V D G D T

FIGURE 4D


10
20
30
SEQ ID NO 186
SEQ ID NO 1958
SEQ ID NO 1960
SEQ ID NO 1962
SEQ ID NO 1238
SEQ ID NO 1242
SEQ ID NO 1240
SEQ ID NO 1241
SEQ ID NO 1243
SEQ ID NO 1222
SEQ ID NO 1223
SEQ ID NO 1232
SEQ ID NO 1221
SEQ ID NO 1231
SEQ ID NO 1227
SEQ ID NO 1235
SEQ ID NO 1230
SEQ ID NO 1229
SEQ ID NO 1228
SEQ ID NO 1246
SEQ ID NO 1247
SEQ ID NO 1244
SEQ ID NO 1245

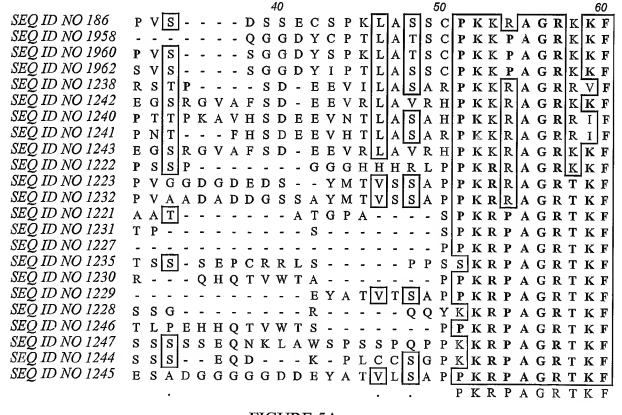

40
50
60
SEQ ID NO 186
SEQ ID NO 1958
SEQ ID NO 1960
SEQ ID NO 1962
SEQ ID NO 1238
SEQ ID NO 1242
SEQ ID NO 1240
SEQ ID NO 1241
SEQ ID NO 1243
SEQ ID NO 1222
SEQ ID NO 1223
SEQ ID NO 1232
SEQ ID NO 1221
SEQ ID NO 1231
SEQ ID NO 1227
SEQ ID NO 1235
SEQ ID NO 1230
SEQ ID NO 1229
SEQ ID NO 1228
SEQ ID NO 1246
SEQ ID NO 1247
SEQ ID NO 1244
SEQ ID NO 1245

FIGURE 5A

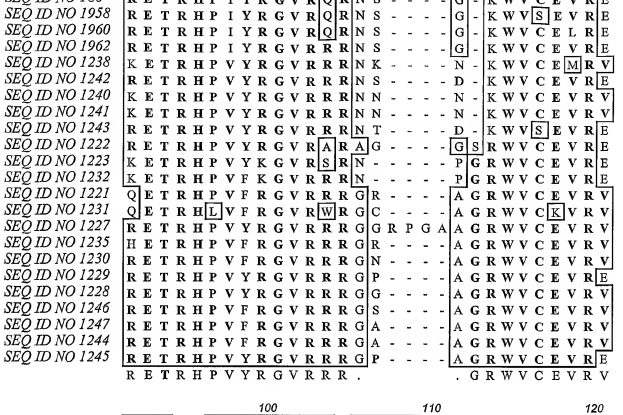

70
80
90
SEQ ID NO 186 R E T R H P I Y R G V R Q R N S - - - - G - K W V C E V R E
SEQ ID NO 1958 R E T R H P I Y R G V R Q R N S - - - - G - K W V S E V R E
SEQ ID NO 1960 R E T R H P I Y R G V R Q R N S - - - - G - K W V C E L R E
SEQ ID NO 1962 R E T R H P I Y R G V R R R N S - - - - G - K W V C E V R E
SEQ ID NO 1238 K E T R H P V Y R G V R R R N K - - - - N - K W V C E M R V
SEQ ID NO 1242 R E T R H P V Y R G V R R R N S - - - - D - K W V C E V R E
SEQ ID NO 1240 K E T R H P V Y R G V R R R N N - - - - N - K W V C E V R V
SEQ ID NO 1241 K E T R H P V Y R G V R R R N N - - - - N - K W V C E V R V
SEQ ID NO 1243 R E T R H P V Y R G V R R R N T - - - - D - K W V S E V R E
SEQ ID NO 1222 R E T R H P V Y R G V R A R A G - - - - G S R W V C E V R E
SEQ ID NO 1223 K E T R H P V Y K G V R S R N - - - - - P G R W V C E V R E
SEQ ID NO 1232 K E T R H P V F K G V R R R N - - - - - P G R W V C E V R E
SEQ ID NO 1221 Q E T R H P V F R G V R R R G R - - - - - A G R W V C E V R V
SEQ ID NO 1231 Q E T R H L V F R G V R W R G C - - - - A G R W V C K V R V
SEQ ID NO 1227 R E T R H P V Y R G V R R R G G R P G A A G R W V C E V R V
SEQ ID NO 1235 H E T R H P V F R G V R R R G R - - - - A G R W V C E V R V
SEQ ID NO 1230 R E T R H P V F R G V R R R G N - - - - A G R W V C E V R V
SEQ ID NO 1229 R E T R H P V Y R G V R R R G P - - - - A G R W V C E V R E
SEQ ID NO 1228 R E T R H P V Y R G V R R R G G - - - - A G R W V C E V R V
SEQ ID NO 1246 R E T R H P V F R G V R R R G S - - - - A G R W V C E V R V
SEQ ID NO 1247 R E T R H P V F R G V R R R G A - - - - A G R W V C E V R V
SEQ ID NO 1244 R E T R H P V F R G V R R R G A - - - - A G R W V C E V R V
SEQ ID NO 1245 R E T R H P V Y R G V R R R G P - - - - A G R W V C E V R E
R E T R H P V Y R G V R R R . . G R W V C E V R V
100
110
120
SEQ ID NO 186 P N K - K S R I W L G T F P T V E M A A R A H D V A A L A L
SEQ ID NO 1958 P N K - K T R I W L G T F Q T A E M A A R A H D V A A L A L
SEQ ID NO 1960 P N K - K T R I W L G T F Q T A E M A A R A H D V A A I A L
SEQ ID NO 1962 P N K - K T R I W L G T F Q T A E M A A R A H D V A A L A L
SEQ ID NO 1238 P N N - N S R I W L G T Y P T P E M A A R A H D V A A L A L
SEQ ID NO 1242 P N K - K T R I W L G T F P T P E M A A R A H D V A A M A L
SEQ ID NO 1240 P N D K S T R I W L G T Y P T P E M A A R A H D V A A L S L
SEQ ID NO 1241 P N D K S T R I W L G T Y P V P E M A A R A H D V A A L A L
SEQ ID NO 1243 P N K - K T R I W L G T F P T P E M A A R A H D V A A M A L
SEQ ID NO 1222 P Q - A Q A R I W L G T Y P T P E M A A R A H D V A A I A L
SEQ ID NO 1223 P H G - K Q R I W L G T F E T A E M A A R A H D V A A M A L
SEQ ID NO 1232 P H G - K Q R I W L G T F E T A E M A A R A H D V A A L A L
SEQ ID NO 1221 P G S R G D R L W V G T F D T A E E A A R A H D A A M L A L
SEQ ID NO 1231 P G S R G D R F W I G T S D T A E E T A R T H D A A M L A L
SEQ ID NO 1227 P G A R G S R L W L G T F A T A E A A A R A H D A A A L A L
SEQ ID NO 1235 P G R R G C R L W L G T F D A A D A A A R A H D A A M L A L
SEQ ID NO 1230 P G R R G C R L W L G T F D T A E G A A R A H D A A M L A I
SEQ ID NO 1229 P N K - K S R I W L G T F A T A E A A A R A H D V A A L A L
SEQ ID NO 1228 P G K R G A R L W L G T Y V T A E A A A R A H D A A M I A L
SEQ ID NO 1246 P G R R G C R L W L G T F D T A E A A A R A H D A A M L A L
SEQ ID NO 1247 P G R R G A R L W L G T Y L G A E A A A R A H D A A M L A L
SEQ ID NO 1244 P G R R G A R L W L G T Y L G A E A A A R A H D A A M L A L
SEQ ID NO 1245 P N K - K S R I W L G T F A T P E A A A R A H D V A A L A L
P . . . R I W L G T F T A E M A A R A H D V A A L A L

FIGURE 5B

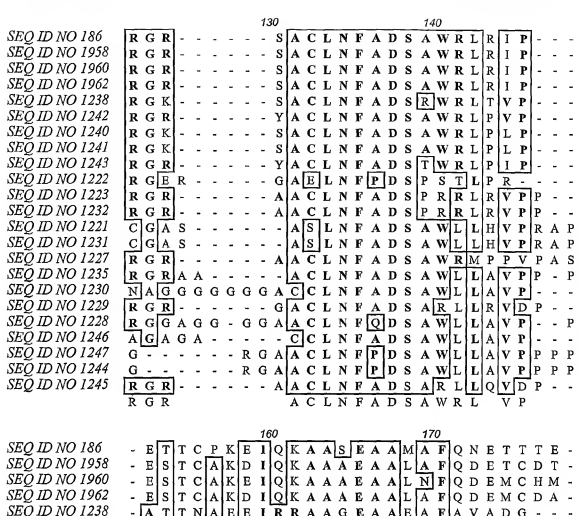

130
140
150
160
170
180
SEQ ID NO 186
SEQ ID NO 1958
SEQ ID NO 1960
SEQ ID NO 1962
SEQ ID NO 1238
SEQ ID NO 1242
SEQ ID NO 1240
SEQ ID NO 1241
SEQ ID NO 1243
SEQ ID NO 1222
SEQ ID NO 1223
SEQ ID NO 1232
SEQ ID NO 1221
SEQ ID NO 1231
SEQ ID NO 1227
SEQ ID NO 1235
SEQ ID NO 1230
SEQ ID NO 1229
SEQ ID NO 1228
SEQ ID NO 1246
SEQ ID NO 1247
SEQ ID NO 1244
SEQ ID NO 1245

FIGURE 5C

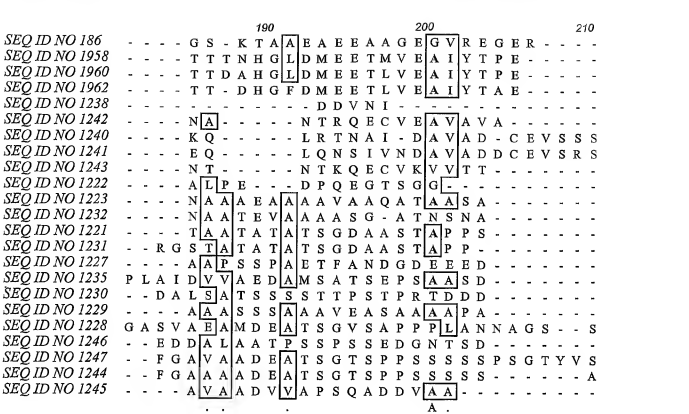

190
200
210
220
230
240
SEQ ID NO 186
SEQ ID NO 1958
SEQ ID NO 1960
SEQ ID NO 1962
SEQ ID NO 1238
SEQ ID NO 1242
SEQ ID NO 1240
SEQ ID NO 1241
SEQ ID NO 1243
SEQ ID NO 1222
SEQ ID NO 1223
SEQ ID NO 1232
SEQ ID NO 1221
SEQ ID NO 1231
SEQ ID NO 1227
SEQ ID NO 1235
SEQ ID NO 1230
SEQ ID NO 1229
SEQ ID NO 1228
SEQ ID NO 1246
SEQ ID NO 1247
SEQ ID NO 1244
SEQ ID NO 1245

FIGURE 5D

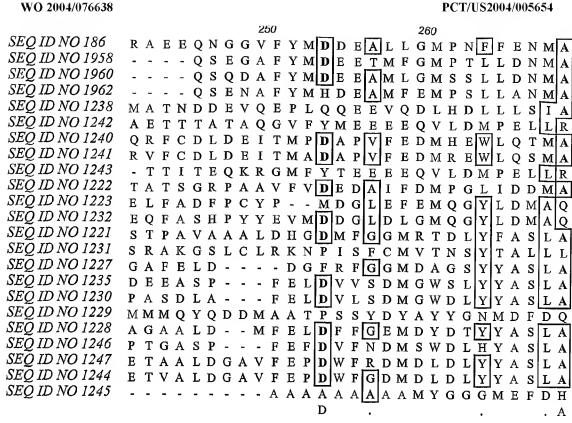

250
260
270
SEQ ID NO 1986
SEQ ID NO 1958
SEQ ID NO 1960
SEQ ID NO 1962
SEQ ID NO 1238
SEQ ID NO 1242
SEQ ID NO 1240
SEQ ID NO 1241
SEQ ID NO 1243
SEQ ID NO 1222
SEQ ID NO 1223
SEQ ID NO 1232
SEQ ID NO 1221
SEQ ID NO 1231
SEQ ID NO 1227
SEQ ID NO 1235
SEQ ID NO 1230
SEQ ID NO 1229
SEQ ID NO 1228
SEQ ID NO 1246
SEQ ID NO 1247
SEQ ID NO 1244
SEQ ID NO 1245
280
290
300
SEQ ID NO 1986
SEQ ID NO 1958
SEQ ID NO 1960
SEQ ID NO 1962
SEQ ID NO 1238
SEQ ID NO 1242
SEQ ID NO 1240
SEQ ID NO 1241
SEQ ID NO 1243
SEQ ID NO 1222
SEQ ID NO 1223
SEQ ID NO 1232
SEQ ID NO 1221
SEQ ID NO 1231
SEQ ID NO 1227
SEQ ID NO 1235
SEQ ID NO 1230
SEQ ID NO 1229
SEQ ID NO 1228
SEQ ID NO 1246
SEQ ID NO 1247
SEQ ID NO 1244
SEQ ID NO 1245

FIGURE 5E

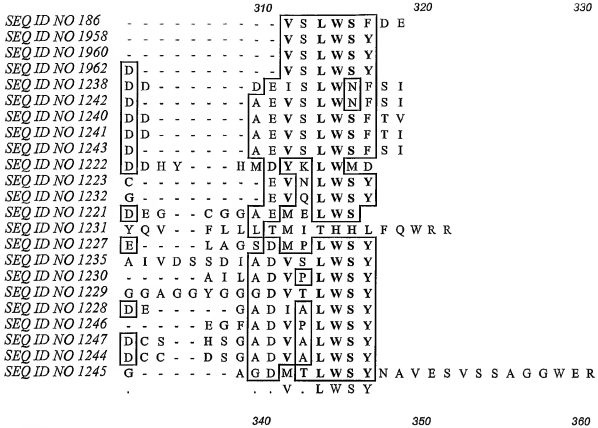

310
320
330
SEQ ID NO 186
SEQ ID NO 1958
SEQ ID NO 1960
SEQ ID NO 1962
SEQ ID NO 1238
SEQ ID NO 1242
SEQ ID NO 1240
SEQ ID NO 1241
SEQ ID NO 1243
SEQ ID NO 1222
SEQ ID NO 1223
SEQ ID NO 1232
SEQ ID NO 1221
SEQ ID NO 1231
SEQ ID NO 1227
SEQ ID NO 1235
SEQ ID NO 1230
SEQ ID NO 1229
SEQ ID NO 1228
SEQ ID NO 1246
SEQ ID NO 1247
SEQ ID NO 1244
SEQ ID NO 1245
340
350
360
SEQ ID NO 186
SEQ ID NO 1958
SEQ ID NO 1960
SEQ ID NO 1962
SEQ ID NO 1238
SEQ ID NO 1242
SEQ ID NO 1240
SEQ ID NO 1241
SEQ ID NO 1243
SEQ ID NO 1222
SEQ ID NO 1223
SEQ ID NO 1232
SEQ ID NO 1221
SEQ ID NO 1231
SEQ ID NO 1227
SEQ ID NO 1235
SEQ ID NO 1230
SEQ ID NO 1229
SEQ ID NO 1228
SEQ ID NO 1246
SEQ ID NO 1247
SEQ ID NO 1244
SEQ ID NO 1245

FIGURE 5F

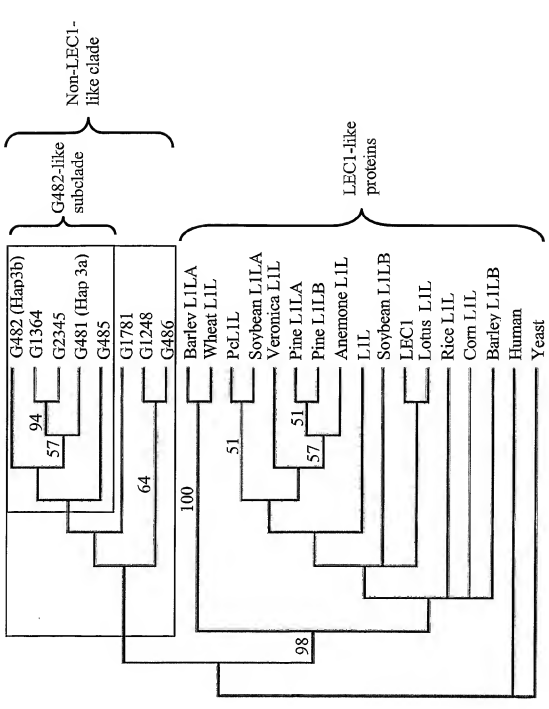

Non-LEC1-like clade
G482-like subclade
LEC1-like proteins
G482 (Hap3b)
G1364
G2345
G481 (Hap 3a)
G485
G1781
G1248
G486
Barley L1LA
Wheat L1L
PeL1L
Soybean L1LA
Veronica L1L
Pine L1LA
Pine L1LB
Anemone L1L
L1L
Soybean L1LB
LEC1
Lotus L1L
Rice L1L
Corn L1L
Barley L1LB
Human
Yeast
94
57
64
100
51
51
57
98

FIGURE 6

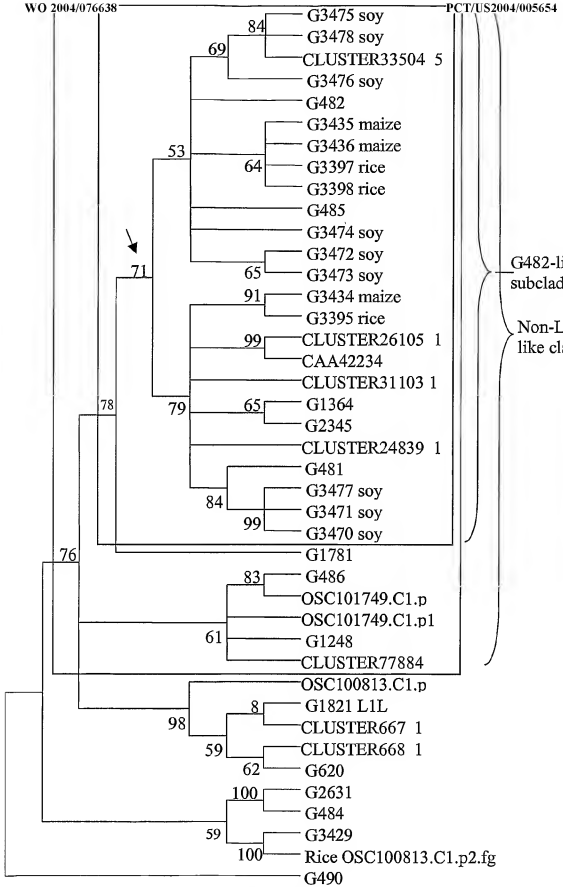

G3475 soy
G3478 soy
CLUSTER33504 5
G3476 soy
G482
G3435 maize
G3436 maize
G3397 rice
G3398 rice
G485
G3474 soy
G3472 soy
G3473 soy
G3434 maize
G3395 rice
CLUSTER26105 1
CAA42234
CLUSTER31103 1
G1364
G2345
CLUSTER24839 1
G481
G3477 soy
G3471 soy
G3470 soy
G1781
G486
OSC101749.C1.p
OSC101749.C1.p1
G1248
CLUSTER77884
OSC100813.C1.p
G1821 L1L
CLUSTER667 1
CLUSTER668 1
G620
G2631
G484
G3429
Rice OSC100813.C1.p2.fg
G490
84
69
53
64
71
65
91
99
78
79
65
84
99
76
83
61
8
98
59
62
100
59
100
G482-like subclade
Non-LEC1-like clade

FIGURE 7

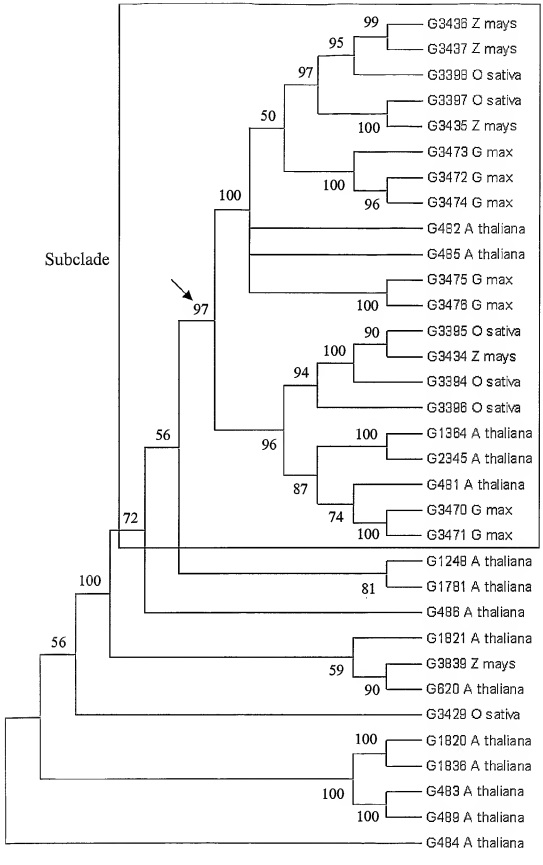

Subclade
99
95
97
50
100
100
96
100
97
100
90
100
94
56
96
100
87
74
100
72
100
81
56
59
90
100
100
100
G3438 Z mays
G3437 Z mays
G3398 O sativa
G3397 O sativa
G3435 Z mays
G3473 G max
G3472 G max
G3474 G max
G482 A thaliana
G485 A thaliana
G3475 G max
G3476 G max
G3395 O sativa
G3434 Z mays
G3394 O sativa
G3396 O sativa
G1364 A thaliana
G2345 A thaliana
G481 A thaliana
G3470 G max
G3471 G max
G1248 A thaliana
G1781 A thaliana
G486 A thaliana
G1821 A thaliana
G3839 Z mays
G620 A thaliana
G3429 O sativa
G1820 A thaliana
G1836 A thaliana
G483 A thaliana
G489 A thaliana
G484 A thaliana

FIGURE 8

FIGURE 9

```
G3472       ----------MABS----------DNESGGHTGNASGSN----------
G3473       ----------MADS----------DNDSGGAHNGGKGS----------
G3474       ----------MABS----------DNESGGHTGNASGSN----------
G3435       ----------MPDS----------DNDSGGPSN-AGG-----------
G3397       ----------MPDS----------DNDSGGPSNYAGG-----------
G3436       ----------MPDS----------DNESGGPSN--A------------
G3398       ----------MPDS----------DNESGGPSN--AG-----------
G3475       ----------MADS----------DNDSGGAHNAGKG-----------
G3478       ----------MADS----------DNDSGGAHNGGKG-----------
G3476       ----------MABS----------DNDSGGAQNAGNSGNL--------
G485        ----------MADS----------DNDSGGHKDGGN------------
G482        ----------MGDS----------DRDSGGQNGNNQNGQ----------
OSC30077    MKSRKSYGHLLSPVGSPPL-----------DNESGEAAAAAAAGGGCGSSAGYVVYGG
G3471       ----------MSDAPPSP----------THESGGEQSPRGSS--------
G3477       ----------MSDALPASP----------SHESGGEQSPRGSL--------
G3470       ----------MSDAPASP-----------SHESGGEQSPRGSL--------
G481        ----------MADTPSSP-----------AGD-GGESG-------------
G1364       ----------MAESQAKSP---GGCGSHESGGDQSPR-------------
G2345       ----------MAESQTGG---GGGGSHESGGDQSPR--------------
AP004366    ----------MADA----------GHDESGSPPRSGGVR-----------
G3434       ----------MADD----------GGSHEGSG-GGGVR------------
G3394       ----------MADGPGSPG---GGGGSHESGSPRGGGGGGGG--------
G1781       ----------MTESSPEEDHGSPGYAETNFGSPSSKTNNNNN--------
G1248       ----------MAGNYHSFQNPIPRYQNYNFGSSSSNHQHEHDGLVV----VVEDQ
G486        ----------------------------------------------------
G1821       ----------MAEGSMEPP----------EFNQPNKTSNGGEE--------
Consensus                                 GG
```

FIGURE 10A

DNA binding domain Subunit interaction →

G3472	---EFSGCREQDRFLPIANVSRIMKKALPANAKISKDAKETVQECVSEFISFTT-GEASD
G3473	---EMS-PREQDRFLPIANVSRIMKKALPANAKISKDAKETVQECVSEFISFHSPGGLAG
G3474	---ELSGCREQDRFLPIANVSRIMKKALPANAKISKEAKETVQECVSEFISFTT-GEASD
G3435	---ELSSPREQDRFLPIANVSRIMKKALPANAKISKDAKETVQECVSEFISFTT-GEASD
G3397	---ELSSPREQDRFLPIANVSRIMKKALPANAKISKDAKETVQECVSEFISFTT-GEASD
G3436	---EFSSPREQDRFLPIANVSRIMKKALPANAKISKDAKETVQECVSEFISFTT-GEASD
G3398	---EYASAREQDRFLPIANVSRIMKRALPANAKISKDAKETVQECVSEFISFTT-GEASD
G3475	---SEMSPREQDRFLPIANVSRIMKKALPANAKISKDAKETVQECVSEFISFTT-GEASD
G3478	---SEMSPREQDRFLPIANVSRIMKKALPANAKISKDAKETVQECVSEFISFTT-GEASD
G3476	---SELSPREQDRFLPIANVSRIMKKALPANAKISKDAKETVQECVSEFISFTT-GEASD
G485	-----ASTREQDRFLPIANVSRIMKKALPANAKISKDAKETVQECVSEFISFTT-GEASD
G482	---SSLSPREQDRFLPIANVSRIMKKALPANAKISKDAKETMQECVSEFISFVT-GEASD
OSC30077	GGGGDSPAKEQDRFLPIANVSRIMKRSLPANAKISKESKETVQECVSEFISFVT-GEASD
G3471	----SG-AREQDRYLPIANISRIMKKALPPNGKIAKDAKDTMQECVSEFISFTT-SEASE
G3477	----SGAAREQDRYLPIANISRIMKKALPPNGKIAKDAKDTMQECVSEFISFTT-SEASE
G3470	-----SGAAREQDRYLPIANISRIMKKALPPNGKIGKDAKDTVQECVSEFISFTT-SEASD
G481	-----GSVREQDRYLPIANISRIMKKALPPNGKIGKDAKDTVQECVSEFISFTT-SEASD
G1364	----SLHVREQDRFLPIANISRIMRRGLPANGKIAKDAKEIVQECVSEFISFVT-SEASD
G2345	----SLNVREQDRFLPIANISRIMRRGLPLNGKIAKDAKETMQECVSEFISFVT-SEASD
AP004366	----------EQDRFLPIANISRIMKKAVPANGKIAKDAKETLQECVSEFISFVT-SEASD
G3434	----------EQDRFLPIANISRIMKKAVPANGKIAKDAKETLQECVSEFISFVT-SEASD
G3394	--GGGGPLVRQDRFLPIANISRIMKKAIPANGKIAKDAKETVQECVSEFISFTT-SEASD
G1781	------NNKEQDRFLPIANVGRIMKKVLPGNGKISKDAKETVQECVSEFISFVT-GEASD
G1248	QQEESMMVKEQDRILLPIANVGRIMKNILPANAKVSKEAKETMQECVSEFISFVT-GEASD
G486	-------MTDEDRILLPIANVGRLMKQILPSNAKLSKEAKQTVQECATEFISFVT-CEASE
G1821	----ECTVREQDRFMPIANVLRIMRRILPAHAKISDDSKETIQECVSEYISFTT-GEANE
Consensus	EQDRFLPIANVSRIMK ALPANAKISKDAKETVQECVSEFISFTT GEASD

FIGURE 10B ← B domain

```
G3472          KCQKEKRKTTNGDDLLWAMTTLGFEEVVEPLKVTLHKYRELEGEKTAMMG----------
G3473          ECQKEKRKTTNGDDLLWAMTTLGFEEVVEPLKVTLHKYRELEGEKTAMMG----------
G3474          KCQKEKRKTTNGDDLLWAMTTLGFEDVVDPLKITLHKYREMEGEKTAMMG----------
G3435          KCQREKRKTTNGDDLLWAMTTLGFEDVVEPLKHTLHKFREIEGERAAASAGASGSQQQ--
G3397          KCQREKRKTTNGDDLLWAMTTLGFEDVVDPLKHTLHKFREIEGERAAASTTGAGTSAA--
G3436          KCQREKRKTTNGDDLLWAMTTLGFEDVVEPLKLYLHKFRELEGEKAANTTSASSGPQPPLH
G3398          KCQREKRKTTNGDDLLWAMTTLGFEDYIDPLKLYLHKFRELEGEKAIGAAGSGGGGAASS
G3475          KCQREKRKTTNGDDLLWAMTTLGFEDVVEPLKGYLQRFREMEGEKTVAAR----------
G3478          KCQREKRKTTNGDDLLWAMTTLGFEDVVEPLKGYLQRFREMEGEKTVAAR----------
G3476          KCQREKRKTTNGDDLLWAMTTLGFEEVVEPLKIYLQRFREMEGEKTVAAR----------
G485           KCQREKRKTTNGDDLLWAMTTLGFEDVVEPLKVYLQKYREVEGEKTTTAGR----------
G482           KCQKEKRKTTNGDDLLWAMTTLGFEDVVEPLKVYLQRFREIEGERTGLGRP----------
OSC30077       KCQREKRKTTNGDDLLWAMTTLGFEAVVGPLKSYLNRYREAEGEKADVLGGAGGAAAARH
G3471          KCQKEKRKTTNGDDLLWAMATLGFEDYIEPLKVYLARYREAEGDTKGSARS----------
G3477          KCQKEKRKTTNGDDLLWAMATLGFEDYIEPLKVTLARYREAEGDTKGSARS----------
G3470          KCQKEKRKTTNGDDLLWAMATLGFEDYIEPLKVYLARYREAEGDTKGSARS----------
G481           KCQKEKRKTVNGDDLLWAMATLGFEDYLEPLKIYLARYRELEGDNKGSGKS----------
G1364          KCQREKRKTTNGDDLLWAMATLGFEDYMEPLKVYLMRYRE--GDTKGSAKG----------
G2345          KCQREKRKTTNGDDLLWAMATLGFEDYIDPLKVYLMRYREMEGDTKGSGKG----------
AP0004366      KCQKEKRKTTNGEDLLFAMGTLGFEEYVDPLKIYLHKYREMEGDSKLSSKA----------
G3434          KCQKEKRKTTNGDDLLWAMATLGFEEYVEPLKIYLQKYKEMEGDSKLSTKA----------
G3394          KCQREKRKTTNGDDLLWAMATLGFEDYIEPLKVTLQKYREMEGDSKLTAKA----------
G1781          KCQREKRKTTNGDDIIWATTTLGFEDVVAPLKVYLCKYRDTEGEKVNSPKQ----------
G1248          KCHKEKRKTTVNGDDICWAMANLGFDDYAAQLKKYLHRYRVLEGEKPN-------------
G486           KCHRENRKTTVNGDDIWWALSTLGLDNYADAVGRHLHKYREAERETEHNKG----------
G1821          RCQREQRKTTTAEDVLWAMSKLGFDDYIEPLTLYLHRYRELEGERGVSCSAG----------
Consensus      KCQREKRKTTNGDDLLWAMTTLGFEDY EPLKVYL YRE EGE
```

B domain

FIGURE 10C

```
G3472      ----------RPHERDEGYGH----------ATPMM-IMMGHQ----------
G3473      ----------RPHERDEGYGH----------ATPMM-IMMGHQ----------
G3474      ----------RPHERDEGYGHGHG--------HATPMMIMMMGHQPQHQHQH--
G3435      --QQQGELPRGAANAAG-YAGYGAPG-------SG-GMMMMMMGQPMYGGSQPQ-
G3397      --STTPQQQHTANAAGGYAGYAAPG-------AGPGGMMMMGQPMYGSPPP--
G3436      RETTPSSSTHN--GAGGPVGGYGMYGGA-----GGSGGMIMMMGQPMYGGSPPA-
G3398      GGSGSGSGSHHHQDASRNNGGYGMYG-------GGGGMIMMMGQPMYG-SPPA-
G3475      --DKDAPPPTNATNSAYESPSYAAA---------PGGIMMHQCHVYG------
G3478      --DKDAPPLTNATNSAYESANYAAAA--------AVPGGIMMHQCHVYG------
G3476      --DSSK---DSASASSY-----------------HQGHVYG------
G485       --QGDKEGGGGGGGAGSGSGGALPMYG-------GGMVTTMGHQFS------
G482       --QTGGEVGEHQRDAVGDGGGFYGGG-------GGMQYHQHHQFL------
OSC30077   GEGGCCGGGGGGGADGVVIDGHYPLAGGLSHSHHGHQQQDGGGDVGLMMGGDAGVGYNAG
G3471      --GDGSATPD-QYGLAGQNSQLVHQG-------SLNYIGLQLQVQP------
G3477      --GDGSATPD-QVGLAGQNSQLVHQG-------SLNYIGLQLQVQP------
G3470      --GDGSARPD-QVGLAGQNAQVQPQX-------SG--YAFNARP------
G481       --GDGSNRDA-GGGVSGEEMPSW-----------------
G1364      --GDPNAKKDGQSSQNGQFSQLAHQG-------PYGNSQ------
G2345      --GESSAKRDGQPSQVSQFSQVPQQG-------SFSQGPYGNSQSLRFGNSIEH
AP004366   --GDGSVKKDTIGPHSCASSSSAQG-------MVGAYTQGMGY------
G3434      --GESVKKDAISPHGGTSSSSNQLV-------QHGVYNQGMGY------
G3394      --GDGSVKKDVLGSHGGSSSSSAQGMG-------QQAAYNQGMGY------
G1781      --QQQRQQQQIQQQNHHNYQFQEQDQNNNN------MSCTSYISHHHPSPFLPVDHQ-
G1248      ----------------------------------HHGKG-----------
G486       --SNDSGNEKE-----------------------TNTRS-----------
G1821      --SVSMTNGLVVKRPNGTMTEYGAYG-------PVPGIHMAQYHYR------
```

FIGURE 10D

```
G3472     ----------QQQHQG--HVYGSGTT--------TGSASSARTR----------
G3473     ----------QQQHQG--HVYGSGTT--------TGSASSARTR----------
G3474     ----------QHQHQHQG--HVYGS----------GSASSARTR----------
G3435     --QQPPQPPQQQQQQHQQHHMAMGRGGFGQQ--GGGGGSSSSSGLGRQDRA----
G3397     ----------PPQQQQQQH--HHMAMGRGGFGHHPGGGGGSSSSSGHGRQNRGA--
G3436     --ASSG----------SYPHH---QMAMGGKGAYGYGGGSSSSPS---GLGR----
G3398     --SSAGYAQPPPPHHHHH--QMVMGGKG-AYGHGGGGGGGPSPSSGYGRQDRL----
G3475     ----------SAGFH---QVAGGAIK--------GGPVYPGPGSNAGRPR------
G3478     ----------SAGFH---QVAGGAIK--------GGPAYPGPGSNAGRPR------
G3476     ----------SPAYH---HQVP-----------GPTYPAPG----RPR------
G485      ----------HHFS----------------------------------------
G482      ----------HQQNHMYGATGGGSDS--------GGGAASGRTRT----------
OSC30077  AGSTTTAFYAPAATAASGNKAYCGGDGSRVMEFEGIGGEEESGGGGGGGERGFAGHLHGV
G3471     ----------QHLVMPSMQSHE-----------------------------------
G3477     ----------QHLVMPSMQSHE-----------------------------------
G3470     ----------------------------------------------------------
G481      ----------------------------------------------------------
G1364     ----------VTFFLFSSHSSN----------THH--SLLIC----------
G2345     LEVLMSSTRTLFTITFRDSTMPVVSENLSDPLSIDMDCEAIYHHFIGLLILSCK-----
AP0043366 ----------MQPQSNFHILVVLQSPAFPYMYQVAQIYCKYPSIE----
G3434     ----------MQPQ----------YHNGET----------
G3394     ----------MQPQ----------YHNGDVSN----------
G1781     ----------PFPNIAFSPKSLQKQ----------FPQQHDNNIDSIHW----------
G1248     ----------GPKSSP----------DN----------
G486      ----------DVQNQSTK----------FIRVVEKGSSSSAR----------
G1821     ----------HQNGFVPSGNEPNSKMSGSSSGASGARVEVFPTQQHKY----
```

FIGURE 10E

```
G3472      ----------
G3473      ----------
G3474      ----------
G3435      ----------
G3397      ----------
G3436      ----------
G3398      ----------
G3475      ----------
G3478      ----------
G3476      ----------
G485       ----------
G482       ----------
OSC30077   QWFRLIKRNTN
G3471      ----------
G3477      ----------
G3470      ----------
G481       ----------
G1364      ----------
G2345      ----------
AP004366   ----------
G3434      ----------
G3394      ----------
G1781      ----------
G1248      ----------
G486       ----------
G1821      ----------
```

FIGURE 10F

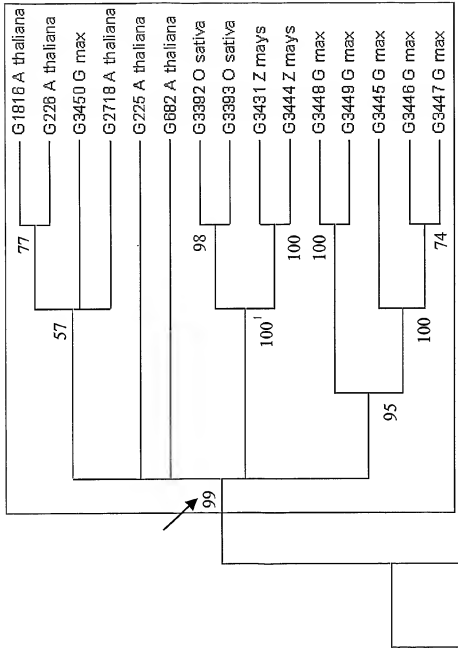

G1816 A thaliana
G228 A thaliana
G3450 G max
G2718 A thaliana
G225 A thaliana
G682 A thaliana
G3382 O sativa
G3383 O sativa
G3431 Z mays
G3444 Z mays
G3448 G max
G3449 G max
G3445 G max
G3446 G max
G3447 G max
G1332 A thaliana
G676 A thaliana
G211 A thaliana
G669 A thaliana
G663 A thaliana
Subclade
77
57
98
100
100
100
95
100
74
54
99

FIGURE 11

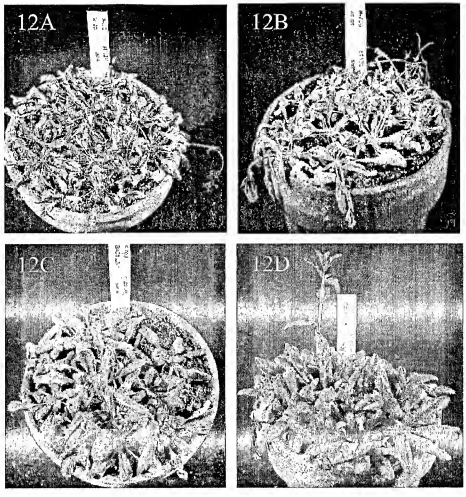


FIGURE 12

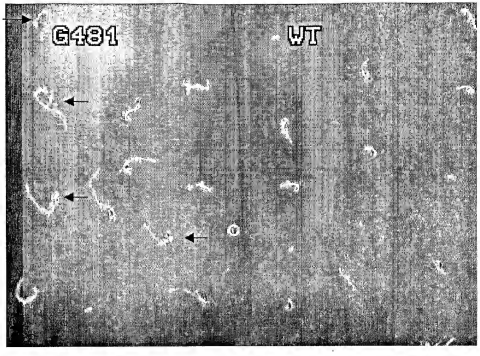


FIGURE 13A

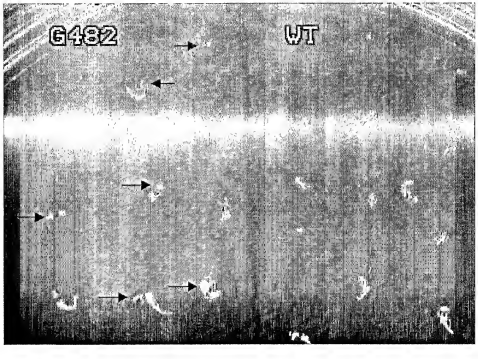


FIGURE 13B

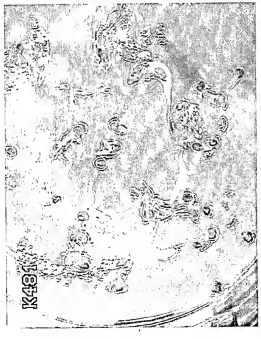


FIGURE 14A

FIGURE 14B

FIGURE 14C

FIGURE 14D

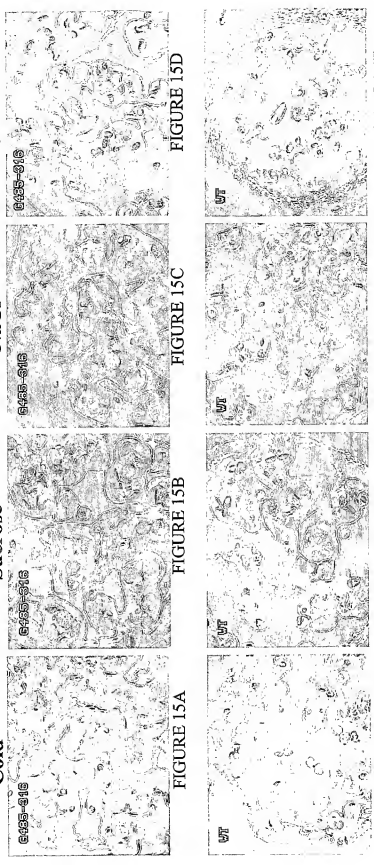


FIGURE 15A FIGURE 15B FIGURE 15C FIGURE 15D

FIGURE 15E FIGURE 15F FIGURE 15G FIGURE 15H

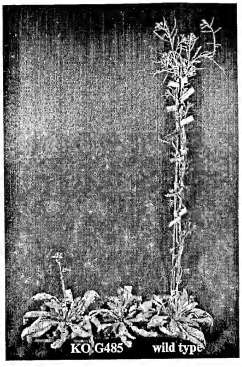


FIGURE 16A

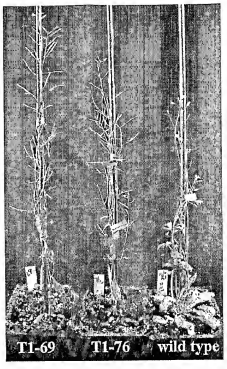


FIGURE 16B

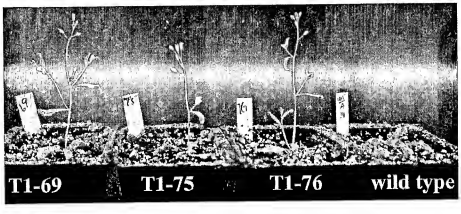


FIGURE 16C

FIGURE 17A

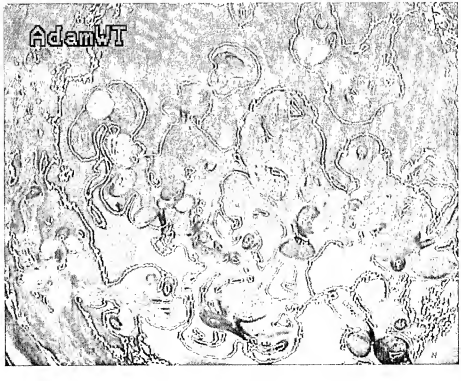


FIGURE 17B

FIGURE 18A

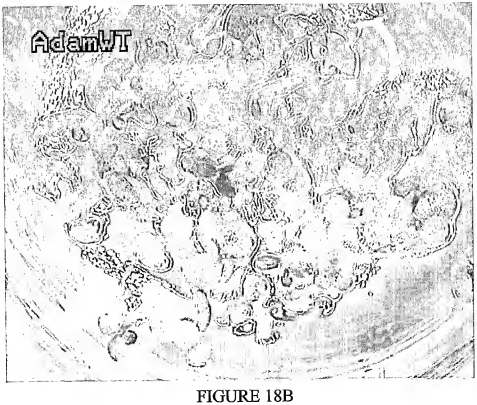


FIGURE 18B

FIGURE 19A

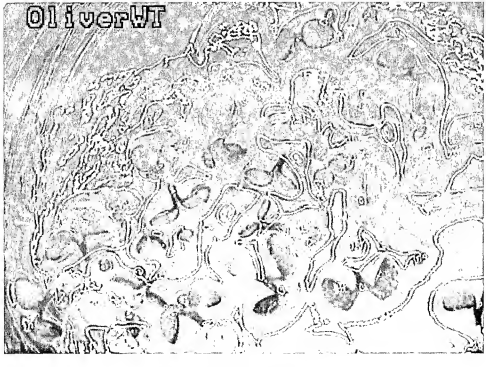


FIGURE 19B

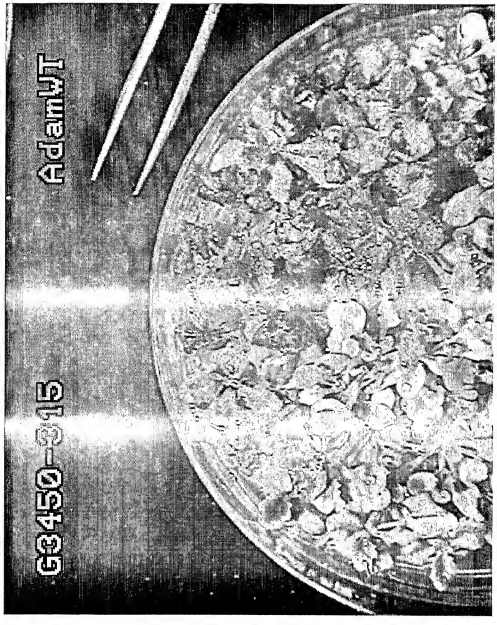


FIGURE 20